A HUNT FOR BACTERIAL GENETIC DETERMINANTS OF EXPEC HOST COLONIZATION

by

Colin W. Russell

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Microbiology and Immunology

Department of Pathology

The University of Utah

December 2016

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Colin W. Russell**

has been approved by the following supervisory committee members:

Matthew Mulvey	, Chair	**10/17/2016** Date Approved
June Round	, Member	**10/17/2016** Date Approved
Kael Fischer	, Member	**10/17/2016** Date Approved
Adam Barker	, Member	**10/17/2016** Date Approved
Joel Griffitts	, Member	**10/17/2016** Date Approved

and by **Peter Jensen**, Chair/Dean of the Department/College/School of **Pathology**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Extraintestinal pathogenic *Escherichia coli* (ExPEC) is a subset of *E. coli* that has evolved the ability to successfully colonize multiple host niches. It spends most of its time as a commensal within the gut as a frequent member of the microbiota. When ExPEC disperses to extraintestinal sites, it can cause diseases such as urinary tract infections (UTIs), bacteremia, and neonatal meningitis. The cumulative impact of ExPEC infections is quite high given the sheer prevalence of UTIs and the severity of bacteremia and neonatal meningitis. A deeper understanding of ExPEC host colonization could lead to innovative therapies that would reduce the disease and financial burden imposed by this pathogen. Here, we investigate the bacterial genetics that underpin host colonization, searching for ExPEC genes that are required for persistence in the gut and infection of extraintestinal sites. We employed three separate approaches, each uncovering a bacterial function critical for host colonization. (1) We studied how *rqlI* contributes to bacterial fitness *in vivo*, and found that in the absence of RqlI, the activity of the helicase RqlH is toxic, creating DNA lesions that trigger the bacterial DNA repair response, inhibit growth, and induce mutations. We propose that RqlH and RqlI function together to modify or repair DNA. (2) A high-throughput screen was employed to find ExPEC genes necessary for metabolism of intestinal mucus, which is thought to provide nutrients to commensal *E. coli* in the gut. The beta-oxidation pathway, which breaks down fatty acids, was found to be an important aspect of mucus metabolism. The beta-oxidation pathway was disrupted by misregulation of the glycerol degradation pathway, as observed in the *glpG* mutant, a strain that exhibited a defect in gut

colonization. (3) To see if canonical extraintestinal virulence factors are important for gut colonization, several mutants were created and tested *in vivo*. A mutant in type 1 pili had reduced fitness, which was only partly attributable to an increase in flagellar expression. Hopefully our work, and similar future endeavors to understand the mechanisms of ExPEC colonization, will improve current therapies and create new treatment strategies.

“At a time when microbiologic research has gained us so many laurels by following the research methods of Koch into the regions of the etiology and pathology of infectious diseases, it would appear to be a pointless and doubtful exercise to examine and disentangle the apparently randomly appearing bacteria in normal feces and the intestinal tract, a situation that seems controlled by a thousand coincidences.”

- Theodor Escherich, 1885, in Fortschritte der Medicin

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I am indebted to my mentor, Matthew Mulvey, who has always been approachable and understanding, and who has taught me almost everything I know about the practice of science. I have been fortunate to be surrounded by wonderful colleagues in the Mulvey Lab, in an environment full of creativity and energy. The Mulvey Lab is a microcosm of the Pathology Department as a whole, which has always focused my thinking and given me encouragement during my presentations and through informal conversations. Many of these conversations were with the members of my dissertation committee, including Kael Fischer, Adam Barker, June Round, and Joel Griffitts, who have been very helpful when I have asked for assistance and advice. I am grateful to Kael Fischer and Tim Mosbruger who have been willing to entertain long conversations regarding bioinformatics.

CHAPTER 1

INTRODUCTION

The Many Sides of *Escherichia coli*

The German pediatric infectious disease physician Theodor Escherich was the first to describe the bacterium *Escherichia coli*, which he called *Bacterium coli commune* [1, 2]. Escherich observed *E. coli* in infant fecal samples using both a microscope and culturing techniques, and began characterizing its metabolic capabilities. Due to its convenient growth in the laboratory and its genetic tractability, *E. coli* became a critical model organism that was put to use in experiments that shaped our current understanding of molecular biology [3]. Most of this historical work, in addition to much of modern-day cloning, has been performed in nonpathogenic fecal isolates of *E. coli*. However, contained within the *E. coli* species are bacteria that exhibit a wide array of interactions with the host, ranging from commensals to pathogens, in addition to strains that can act as either depending on the context [3]. Several subclasses of *E. coli* induce disease in the gut, and are defined by the type of disease they cause and their mechanism of pathogenesis. For example, enterohemorrhagic *E. coli* (EHEC) can be acquired from contaminated bovine meat, and causes bloody diarrhea and hemolytic uremic syndrome [4]. Another *E. coli* pathotype is extraintestinal pathogenic *E. coli* (ExPEC), a gut commensal that causes disease at extraintestinal sites such as the blood and the urinary tract. The *E. coli* strains that constitute the ExPEC pathotype are the main focus of this work.

The Diseases and Domiciles of ExPEC

ExPEC can cause many types of infections outside of the intestinal tract, the most prevalent of which is the urinary tract infection (UTI). ExPEC infection of the urinary tract occurs in an ascending manner, with the bacteria climbing the urethra to the bladder, which is sometimes followed by transit up the ureters to the kidneys [5]. Bladder infection, known as cystitis, causes many symptoms, including painful

urination, suprapubic pain, a frequent and urgent feeling of needing to urinate, and cloudy urine [6]. Bladder infections are sometimes accompanied with infection of the kidneys, known as pyelonephritis, the symptoms of which include back pain, vomiting, fever, and nausea, in addition to the symptoms associated with cystitis [6]. UTIs are typically addressed in the outpatient setting, with a diagnosis that is based on symptoms and medical history, and that is confirmed with urinalysis. Treatment involves a short course of antibiotics to clear the infection. UTIs can either be acquired in the community or contracted during a hospital stay, most often from a catheter [7]. The overwhelming majority of community-acquired UTIs are caused by ExPEC, whereas ExPEC is the etiological agent in only ~13% of catheter-associated UTIs [8, 9].

The prevalence of UTIs is quite high, and in the year 2000 in the United States alone, UTIs accounted for over 8 million office visits and 350,000 hospitalizations [10]. UTIs afflict women more often than men, with more than half of all women experiencing a UTI at some point in their lives, whereas the lifetime incidence rate for men is only about 14% [11, 12]. Although the personal economic cost of UTI treatment is often small, the aggregate economic impact of these infections is enormous. It is estimated that per annum, it costs roughly $2.3 billion (2010 USD) to treat community-acquired UTIs, $350 million (2007 USD) for nosocomial UTIs, and $2.14 billion (2000 USD) to treat pyelonephritis [13-15]. It is clear that better preventative measures and new strategies to treat UTIs have the potential to reduce not only the disease burden, but also the detrimental economic effects.

ExPEC are also capable of infecting the bloodstream where they cause more serious symptoms, including death. ExPEC is the most common cause of bloodstream infections worldwide, followed by *Staphylococcus aureus* and *Streptococcus pneumoniae* [16]. There are many routes through which ExPEC is able to access the blood. For example, newborns can become infected *in utero* or during birth with bacteria

originating from the mother's genitourinary tract [17]. Early onset neonatal sepsis, which occurs during the first 2 days of life, afflicted an average of 165 newborns per 1,000 live births every year from 2005-2008 in the United States, with a mortality rate of ~11%. ExPEC was the second leading etiological agent of neonatal sepsis behind Group B streptococcus, but was the leading cause of death, accounting for ~25% of all mortalities.

In adults, ExPEC can cause urosepsis, which is characterized by bacterial dissemination from the urinary tract to the bloodstream. Evidence that this occurs is given by the observation that usually when a patient suffers from a UTI and bacteremia simultaneously, the same ExPEC strain can be isolated from both the urine and blood [18]. In a large study of ~3,600 bacteremia cases, *E. coli* caused ~24% of episodes, with the most common origin of ExPEC being the urinary tract, leading to a mortality rate of 2.6% [19]. Therefore, although the urinary tract is frequently a source of ExPEC, the mortality rate remains quite low for this type of bacteremia.

Alarmingly, the incidence of bloodstream infections has been increasing over time in some parts of the world, partly due to an ageing population [20]. This older cohort may be especially susceptible to ExPEC accessing the blood from the gut, especially in cases where the gastrointestinal barrier is no longer efficient, as occurs during chemotherapy-induced neutropenia [21, 22]. In cancer patients that suffer from bacteremia, Gram-negative bacteria are often responsible for the infection, with *E. coli* being the most common organism [23, 24]. Furthermore, these *E. coli* isolates often carry ExPEC virulence factors, suggesting that ExPEC is the most common type of *E. coli* to cause disease in bacteremic cancer patients [25]. Evidence that these bacteria originate from the gut is that the infecting strain can be isolated from both the gut and the blood simultaneously [22].

No matter the route or type of extraintestinal infection, the first step appears to usually be gut colonization. Indeed, the gut is thought to be the primary animal niche for

ExPEC, and evidence indicates that many healthy humans are colonized by ExPEC. Much of what we know about ExPEC colonization of the human gut comes from epidemiological studies that analyze the predominant *E. coli* isolate from the feces of volunteers [26]. The analyses involve dividing *E. coli* into 4 main phylogenetic groups, including A, B1, B2, and D [27]. Strains within the B2 phylogroup are enriched with ExPEC virulence factors [28], and over 80% of B2 strains exhibit virulence in a mouse model of septicemia [29]. Therefore, the prevalence of B2 strains within the gut is often assumed to be a rough estimate for ExPEC abundance. In studies that limit the analysis to the predominant fecal *E. coli* isolate, a B2 strain is found to be the major member of the *E. coli* population in almost 30% of healthy subjects on average [26]. This is in agreement with a more refined approach that utilized qPCR to quantify the presence of each of the 4 phylogenetic groups simultaneously in a given sample, and which found B2 to be the dominant group in 37% of cases [30]. B2 *E. coli* was present—dominant or otherwise—in 70% of the individuals [30], suggesting that the majority of healthy humans carry ExPEC within their gut.

The high frequency at which ExPEC is a member of the microbiota may, in part, be due to its ability to persist within the gut. When the phylogroup of the predominant *E. coli* fecal isolate is tracked over time in individuals, resident strains most often belong to the B2 group [31, 32], suggesting that ExPEC may persist within the gut better than other *E. coli* subclasses. However, a similar longitudinal study in a different cohort did not find an association of gut persistence with B2 group strains, but the B2 strains in this study did not carry many ExPEC virulence factors, suggesting that the correlation between B2 and ExPEC may not have held true in this instance [33]. Taken together, these data point to the possibility that ExPEC are able to remain as members of the gut microbiota for extended periods of time.

In addition to being the source of ExPEC for initial infection, the gut is likely a

reservoir for recurring UTIs. Two lines of circumstantial evidence support this idea. First, in patients suffering from recurring UTIs, the same strain frequently causes the initial and subsequent infections [34]. This observation indicates that a reservoir exists that is not sufficiently cleared by antibiotic treatment. Second, as mentioned above, the same ExPEC strain can be isolated from the feces and urine of patients suffering from acute UTI [35-38], suggesting some sort of transit between the two niches. When these observations are considered together, it is clear that the gut may be able to act as a source of ExPEC for both initial and subsequent infections. This hypothesis has not been carefully tested, and raises several questions, including why the gut reservoir would not be cleared by antibiotic treatment during the initial UTI. In addition to the gut, other reservoirs that drive recurring infection could exist, including bacteria residing inside of bladder host cells, or a source of ExPEC located within a sexual partner or a household pet [37, 39, 40].

The Elusive Genetic Definition of ExPEC

An *E. coli* strain belongs to the ExPEC pathogenicity group if it is isolated from an extraintestinal site during infection. ExPEC is defined by its potential for causing extraintestinal disease because it is currently not possible to identify an *E. coli* strain as ExPEC based on its genetic content alone. This is in part due to the wide genetic diversity of the *E. coli* species, which can even surpass the genetic difference between two eukaryotic species. For example, the total divergence between chimpanzees and humans is estimated to be 4-6%, which can mostly be attributed to small indels [41, 42]. In contrast, when the two *E. coli* strains 536 and MG1655 are compared, they exhibit a larger degree of dissimilarity [43]. For example, about ¼ of the genes in 536 have no ortholog in MG1655. This observation can be generalized to the entire species as a whole, and is apparent in the *E. coli* pan-genome, the cumulative genetic content of all *E. coli*

strains. The pan-genome can be divided into two parts, including the core and accessory genomes. The core genome consists of ~3,000 genes that are shared between all *E. coli* strains, whereas the accessory genome includes the genes that are not shared with all other *E. coli* strains [44]. There are an estimated 16,000 genes included in the pan-genome, and this number is expected to rise as more *E. coli* strains are sequenced and new accessory genes are discovered [44].

Because of the wide genetic diversity observed in *E. coli*, and therefore between ExPEC strains, there is no gene or set of genes that appear to define ExPEC. There are several virulence factors that have been associated with ExPEC strains epidemiologically [5], but the association is never perfect, and their role in virulence is not always concretely known. Instead of one common set of virulence factors, it is likely that there are several gene sets that could allow ExPEC to cause pathogenesis, and these sets may vary from one ExPEC strain to the next. There are several genes that are considered to be canonical ExPEC virulence factors, some of which will be examined in this work, including the adhesive organelles type 1 and P pili, the α-hemolysin toxin, and the plasmid pUTI89.

Pili are extracellular organelles that extend from the bacterial outer membrane and mediate adhesion to surfaces via the adhesin protein found at their tip. For type 1 pili, the adhesin protein is FimH, which mediates binding to mannose residues on glycosylated proteins [45]. Type 1 pili are considered a virulence factor due to their ability to mediate various steps in bladder infection, including binding to the uroepithelium, induction of host cell exfoliation, invasion into host cells, and formation of intracellular bacterial communities [39, 46, 47]. Although similar in structure to type 1 pili, P pili constitute a different type of pilus that terminates with PapG, an adhesin that binds to glycosphingolipids found on the surface of host cell membranes [48]. Several studies indicate that P pili may facilitate kidney colonization by binding to the kidney

epithelium [49-51]. Although ExPEC encode and express many types of pili, type 1 and P pili are the best studied, which likely contributes to their status as canonical virulence factors.

The α-hemolysin toxin HlyA is often associated with ExPEC, being present in about half of all UTI isolates, but only 5% of all fecal isolates [52]. The toxin is lipidated and then released via a type 1 secretion system prior to its interaction with host cells [53]. Once adsorbed to host cells, hemolysin can exert many effects, including both cell death and alterations of cell signaling. The full physiological effects of hemolysin during extraintestinal infection are not entirely clear, although it has been associated with virulence in various models [54, 55]. Furthermore, it alters host inflammation through many mechanisms including lysis of leukocytes [56]. Hemolysin may even cause inflammation during gut colonization [57], suggesting that ExPEC may not act as a harmless inhabitant of the microbiota at all times.

Bacterial pathogens often carry and exchange plasmids that encode various virulence factors. The prototypical ExPEC isolate UTI89 propagates a ~114 kb plasmid called pUTI89 that contains the genes required for plasmid transfer, some potential virulence genes, and many genes with unknown functions [58]. This plasmid has since been observed in other ExPEC isolates [59]. When pUTI89 was removed from UTI89, the bacteria exhibited poor fitness within the urinary tract, especially at early time points [60]. Although some specific plasmid-encoded genes that contributed to *in vivo* fitness were implicated, much of the genetic content of pUTI89 remains enigmatic.

There are many other ExPEC genes that have been labeled as virulence factors in addition to type 1 pili, P pili, α-hemolysin, and pUTI89. Comparative genomics can be used to study the origin and evolution of these virulence factors, and the pathogens that encode them. For example, a comparison between ExPEC isolates and non-pathogenic *E. coli* strains helped to define regions that are specific to the ExPEC isolates [61]. Many of

these regions were pathogenicity islands, which are large stretches of DNA that encode virulence factors (such as α-hemolysin). These regions contain hints that they were acquired through horizontal gene transfer from a different species. For example, the DNA in pathogenicity islands frequently have a GC content and codon bias that are uncharacteristic of *E. coli* [62]. As many ExPEC virulence factors can be found within these islands, it appears that the pathogenic nature of ExPEC is, in part, a product of the horizontal acquisition of genetic content from other bacteria.

It is expected that factors newly obtained via horizontal transfer are subsequently more finely molded by the evolutionary pressures that ExPEC experience. Given that most host-associated ExPEC is found within the gut, and that a given ExPEC strain spends most of its history in the gastrointestinal tract, it has been proposed that the gut is the main evolutionary arena for ExPEC [29]. In contrast to the gut, extraintestinal niches such as the urinary tract represent a more transient ExPEC home with a limited transmission frequency. How does an extraintestinal virulence factor evolve, and how is it maintained in the genome, when it is used so infrequently in a pathogenic context? It has been proposed that the hypothesis of coincidental evolution could be applied to the ecology of ExPEC [29]. In general terms, coincidental evolution is when a factor that evolved for survival in one niche happens to be beneficial in another [63]. When applied to ExPEC, the resulting conclusion is that the utility of extraintestinal virulence factors outside of the gut is secondary to their primary function within the gut. Very little evidence exists to support or refute this idea. Many studies have correlated ExPEC virulence factors with persistence within the gut, but a causal link was never established [29, 32, 64]. Another study attempted to demonstrate that pathogenicity islands important for extraintestinal infection also contribute to gut colonization, but it was clear that deletion of the islands led to complicating effects (e.g. increased flagellar expression) that made the results difficult to interpret [65]. Overall, the meshing of

horizontal gene transfer and coincidental evolution forms an attractive model for the evolution of ExPEC virulence factors.

"A Pointless and Doubtful Exercise"

When Theodor Escherich lectured regarding his studies of gut microbes, he highlighted the overwhelming complexity that can be found within the gut microbiota: "…it would appear to be a pointless and doubtful exercise to examine and disentangle the apparently randomly appearing bacteria in normal feces and the intestinal tract, a situation that seems controlled by a thousand coincidences" [1]. Indeed, the gut microbiota is composed of hundreds of bacterial species—in addition to viral and fungal communities—and is influenced by host genetics, diet, and the immune system, in addition to many other factors [66-68]. Here, we take a focused approach to increasing our understanding of this complex ecosystem by endeavoring to define the genes that are important for one member of the microbiota to colonize the gut.

Many strategies are employed in this work to uncover genes important for ExPEC gut colonization. We take a candidate gene approach by querying the role of extraintestinal virulence factors within the gut, effectively testing the hypothesis of coincidental evolution as applied to ExPEC. Additionally, a more unbiased approach is utilized by performing a forward genetic screen to find genes important for metabolism of intestinal mucus. The results from these various strategies begin to paint a picture of the mechanism of ExPEC gut colonization, touching on various processes including DNA maintenance, metabolism, and interactions with the host.

References

1. Escherich T. The intestinal bacteria of the neonate and breast-fed infant. 1884. Rev Infect Dis. 1988;10(6):1220-5. PubMed PMID: 3060950.

2. Shulman ST, Friedmann HC, Sims RH. Theodor Escherich: the first pediatric infectious diseases physician? Clin Infect Dis. 2007;45(8):1025-9. doi:

10.1086/521946. PubMed PMID: 17879920.

3. Blount ZD. The unexhausted potential of E. coli. Elife. 2015;4. doi: 10.7554/eLife.05826. PubMed PMID: 25807083; PubMed Central PMCID: PMCPMC4373459.

4. Kaper JB, Nataro JP, Mobley HL. Pathogenic Escherichia coli. Nat Rev Microbiol. 2004;2(2):123-40. doi: 10.1038/nrmicro818. PubMed PMID: 15040260.

5. Wiles TJ, Kulesus RR, Mulvey MA. Origins and virulence mechanisms of uropathogenic Escherichia coli. Experimental and molecular pathology. 2008;85(1):11-9. doi: 10.1016/j.yexmp.2008.03.007. PubMed PMID: 18482721; PubMed Central PMCID: PMC2595135.

6. Barber AE, Norton JP, Spivak AM, Mulvey MA. Urinary tract infections: current and emerging management strategies. Clin Infect Dis. 2013;57(5):719-24. doi: 10.1093/cid/cit284. PubMed PMID: 23645845; PubMed Central PMCID: PMCPMC3739462.

7. Weber DJ, Sickbert-Bennett EE, Gould CV, Brown VM, Huslage K, Rutala WA. Incidence of catheter-associated and non-catheter-associated urinary tract infections in a healthcare system. Infect Control Hosp Epidemiol. 2011;32(8):822-3. doi: 10.1086/661107. PubMed PMID: 21768769.

8. Hooton TM. Clinical practice. Uncomplicated urinary tract infection. N Engl J Med. 2012;366(11):1028-37. Epub 2012/03/16. doi: 10.1056/NEJMcp1104429. PubMed PMID: 22417256.

9. Tambyah PA, Maki DG. Catheter-associated urinary tract infection is rarely symptomatic: a prospective study of 1,497 catheterized patients. Arch Intern Med. 2000;160(5):678-82. PubMed PMID: 10724054.

10. Litwin MS, Saigal CS, Yano EM, Avila C, Geschwind SA, Hanley JM, et al. Urologic diseases in America Project: analytical methods and principal findings. J Urol. 2005;173(3):933-7. doi: 10.1097/01.ju.0000152365.43125.3b. PubMed PMID: 15711342.

11. Griebling TL. Urologic diseases in america project: trends in resource use for urinary tract infections in men. J Urol. 2005;173(4):1288-94. doi: 10.1097/01.ju.0000155595.98120.8e. PubMed PMID: 15758784.

12. Foxman B, Barlow R, D'Arcy H, Gillespie B, Sobel JD. Urinary tract infection: self-reported incidence and associated costs. Ann Epidemiol. 2000;10(8):509-15. PubMed PMID: 11118930.

13. Foxman B. The epidemiology of urinary tract infection. Nat Rev Urol. 2010;7(12):653-60. doi: 10.1038/nrurol.2010.190. PubMed PMID: 21139641.

14. Scott RDI. The Direct Medical Costs of Healthcare-Associated Infections in U.S. Hospitals and the Benefits of Prevention. In: Prevention CfDCa, editor. cdc.gov2009.

15. Brown P, Ki M, Foxman B. Acute pyelonephritis among adults: cost of illness and

considerations for the economic evaluation of therapy. Pharmacoeconomics. 2005;23(11):1123-42. PubMed PMID: 16277548.

16. Laupland KB. Incidence of bloodstream infection: a review of population-based studies. Clin Microbiol Infect. 2013;19(6):492-500. doi: 10.1111/1469-0691.12144. PubMed PMID: 23398633.

17. Simonsen KA, Anderson-Berry AL, Delair SF, Davies HD. Early-onset neonatal sepsis. Clin Microbiol Rev. 2014;27(1):21-47. doi: 10.1128/CMR.00031-13. PubMed PMID: 24396135; PubMed Central PMCID: PMCPMC3910904.

18. Vollmerhausen TL, Woods JL, Faoagali J, Katouli M. Interactions of uroseptic Escherichia coli with renal (A-498) and gastrointestinal (HT-29) cell lines. J Med Microbiol. 2014;63(Pt 12):1575-83. doi: 10.1099/jmm.0.076562-0. PubMed PMID: 25298161.

19. Gransden WR, Eykyn SJ, Phillips I, Rowe B. Bacteremia due to Escherichia coli: a study of 861 episodes. Rev Infect Dis. 1990;12(6):1008-18. PubMed PMID: 2267481.

20. van der Mee-Marquet NL, Blanc DS, Gbaguidi-Haore H, Dos Santos Borges S, Viboud Q, Bertrand X, et al. Marked increase in incidence for bloodstream infections due to Escherichia coli, a side effect of previous antibiotic therapy in the elderly. Front Microbiol. 2015;6:646. doi: 10.3389/fmicb.2015.00646. PubMed PMID: 26175721; PubMed Central PMCID: PMCPMC4485226.

21. Green SI, Ajami NJ, Ma L, Poole NM, Price RE, Petrosino JF, et al. Murine model of chemotherapy-induced extraintestinal pathogenic Escherichia coli translocation. Infect Immun. 2015;83(8):3243-56. doi: 10.1128/IAI.00684-15. PubMed PMID: 26034214; PubMed Central PMCID: PMCPMC4496622.

22. Samet A, Sledzinska A, Krawczyk B, Hellmann A, Nowicki S, Kur J, et al. Leukemia and risk of recurrent Escherichia coli bacteremia: genotyping implicates E. coli translocation from the colon to the bloodstream. Eur J Clin Microbiol Infect Dis. 2013;32(11):1393-400. doi: 10.1007/s10096-013-1886-9. PubMed PMID: 23649557; PubMed Central PMCID: PMCPMC3824565.

23. Andremont A, Lancar R, An Le N, Hattchouel JM, Baron S, Tavakoli T, et al. Secular trends in mortality associated with bloodstream infections in 4268 patients hospitalized in a cancer referral center between 1975 and 1989. Clin Microbiol Infect. 1996;1(3):160-7. PubMed PMID: 11866751.

24. Tancrede CH, Andremont AO. Bacterial translocation and gram-negative bacteremia in patients with hematological malignancies. J Infect Dis. 1985;152(1):99-103. PubMed PMID: 3925032.

25. Hilali F, Ruimy R, Saulnier P, Barnabe C, Lebouguenec C, Tibayrenc M, et al. Prevalence of virulence genes and clonality in Escherichia coli strains that cause bacteremia in cancer patients. Infect Immun. 2000;68(7):3983-9. PubMed PMID: 10858212; PubMed Central PMCID: PMCPMC101677.

26. Bailey JK, Pinyon JL, Anantham S, Hall RM. Distribution of human commensal Escherichia coli phylogenetic groups. J Clin Microbiol. 2010;48(9):3455-6. doi:

10.1128/JCM.00760-10. PubMed PMID: 20610687; PubMed Central PMCID: PMCPMC2937668.

27. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal Escherichia coli. Nat Rev Microbiol. 2010;8(3):207-17. doi: 10.1038/nrmicro2298. PubMed PMID: 20157339.

28. Johnson JR, Delavari P, Kuskowski M, Stell AL. Phylogenetic distribution of extraintestinal virulence-associated traits in Escherichia coli. J Infect Dis. 2001;183(1):78-88. doi: 10.1086/317656. PubMed PMID: 11106538.

29. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, et al. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group Escherichia coli strains. Mol Biol Evol. 2007;24(11):2373-84. doi: 10.1093/molbev/msm172. PubMed PMID: 17709333.

30. Smati M, Clermont O, Le Gal F, Schichmanoff O, Jaureguy F, Eddi A, et al. Real-time PCR for quantitative analysis of human commensal Escherichia coli populations reveals a high frequency of subdominant phylogroups. Appl Environ Microbiol. 2013;79(16):5005-12. doi: 10.1128/AEM.01423-13. PubMed PMID: 23770894; PubMed Central PMCID: PMCPMC3754702.

31. Nowrouzian FL, Wold AE, Adlerberth I. Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. J Infect Dis. 2005;191(7):1078-83. doi: 10.1086/427996. PubMed PMID: 15747243.

32. Nowrouzian FL, Adlerberth I, Wold AE. Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes Infect. 2006;8(3):834-40. doi: 10.1016/j.micinf.2005.10.011. PubMed PMID: 16483819.

33. Nowrouzian FL, Ostblom AE, Wold AE, Adlerberth I. Phylogenetic group B2 Escherichia coli strains from the bowel microbiota of Pakistani infants carry few virulence genes and lack the capacity for long-term persistence. Clin Microbiol Infect. 2009;15(5):466-72. doi: 10.1111/j.1469-0691.2009.02706.x. PubMed PMID: 19260873.

34. Russo TA, Stapleton A, Wenderoth S, Hooton TM, Stamm WE. Chromosomal restriction fragment length polymorphism analysis of Escherichia coli strains causing recurrent urinary tract infections in young women. The Journal of infectious diseases. 1995;172(2):440-5. Epub 1995/08/01. PubMed PMID: 7622887.

35. Moreno E, Andreu A, Pigrau C, Kuskowski MA, Johnson JR, Prats G. Relationship between Escherichia coli strains causing acute cystitis in women and the fecal E. coli population of the host. J Clin Microbiol. 2008;46(8):2529-34. doi: 10.1128/JCM.00813-08. PubMed PMID: 18495863; PubMed Central PMCID: PMCPMC2519474.

36. Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME, Hultgren SJ, et al. Genomic diversity and fitness of E. coli strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Sci Transl Med.

2013;5(184):184ra60. doi: 10.1126/scitranslmed.3005497. PubMed PMID: 23658245; PubMed Central PMCID: PMCPMC3695744.

37. Foxman B, Manning SD, Tallman P, Bauer R, Zhang L, Koopman JS, et al. Uropathogenic Escherichia coli are more likely than commensal E. coli to be shared between heterosexual sex partners. Am J Epidemiol. 2002;156(12):1133-40. PubMed PMID: 12480658.

38. Yamamoto S, Tsukamoto T, Terai A, Kurazono H, Takeda Y, Yoshida O. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by Escherichia coli. J Urol. 1997;157(3):1127-9. PubMed PMID: 9072556.

39. Mulvey MA, Lopez-Boado YS, Wilson CL, Roth R, Parks WC, Heuser J, et al. Induction and Evasion of Host Defenses by Type 1-Piliated Uropathogenic Escherichia coli. Science. 1998;282(5393):1494-7. doi: 10.1126/science.282.5393.1494.

40. Johnson JR, Clabots C, Kuskowski MA. Multiple-host sharing, long-term persistence, and virulence of Escherichia coli clones from human and animal household members. J Clin Microbiol. 2008;46(12):4078-82. doi: 10.1128/JCM.00980-08. PubMed PMID: 18945846; PubMed Central PMCID: PMCPMC2593269.

41. Wetterbom A, Sevov M, Cavelier L, Bergstrom TF. Comparative genomic analysis of human and chimpanzee indicates a key role for indels in primate evolution. J Mol Evol. 2006;63(5):682-90. doi: 10.1007/s00239-006-0045-7. PubMed PMID: 17075697.

42. Chimpanzee S, Analysis C. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005;437(7055):69-87. doi: 10.1038/nature04072. PubMed PMID: 16136131.

43. Brzuszkiewicz E, Bruggemann H, Liesegang H, Emmerth M, Olschlager T, Nagy G, et al. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic Escherichia coli strains. Proc Natl Acad Sci U S A. 2006;103(34):12879-84. doi: 10.1073/pnas.0603038103. PubMed PMID: 16912116; PubMed Central PMCID: PMCPMC1568941.

44. Kaas RS, Friis C, Ussery DW, Aarestrup FM. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. BMC Genomics. 2012;13:577. doi: 10.1186/1471-2164-13-577. PubMed PMID: 23114024; PubMed Central PMCID: PMCPMC3575317.

45. Krogfelt KA, Bergmans H, Klemm P. Direct evidence that the FimH protein is the mannose-specific adhesin of Escherichia coli type 1 fimbriae. Infect Immun. 1990;58(6):1995-8. PubMed PMID: 1971261; PubMed Central PMCID: PMCPMC258756.

46. Martinez JJ, Mulvey MA, Schilling JD, Pinkner JS, Hultgren SJ. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. Embo J. 2000;19(12):2803-12. Epub 2000/06/17. doi: 10.1093/emboj/19.12.2803. PubMed PMID: 10856226; PubMed Central PMCID: PMC203355.

47. Wright KJ, Seed PC, Hultgren SJ. Development of intracellular bacterial communities of uropathogenic Escherichia coli depends on type 1 pili. Cell Microbiol. 2007;9(9):2230-41. Epub 2007/05/11. doi: 10.1111/j.1462-5822.2007.00952.x. PubMed PMID: 17490405.

48. Leffler H, Svanborg-Eden C. Glycolipid receptors for uropathogenic Escherichia coli on human erythrocytes and uroepithelial cells. Infection and immunity. 1981;34(3):920-9. Epub 1981/12/01. PubMed PMID: 7037645; PubMed Central PMCID: PMC350957.

49. O'Hanley P, Lark D, Falkow S, Schoolnik G. Molecular basis of Escherichia coli colonization of the upper urinary tract in BALB/c mice. Gal-Gal pili immunization prevents Escherichia coli pyelonephritis in the BALB/c mouse model of human pyelonephritis. J Clin Invest. 1985;75(2):347-60. doi: 10.1172/JCI111707. PubMed PMID: 2857730; PubMed Central PMCID: PMCPMC423490.

50. Roberts JA, Hardaway K, Kaack B, Fussell EN, Baskin G. Prevention of pyelonephritis by immunization with P-fimbriae. J Urol. 1984;131(3):602-7. PubMed PMID: 6199527.

51. Wullt B, Bergsten G, Connell H, Rollano P, Gebretsadik N, Hull R, et al. P fimbriae enhance the early establishment of Escherichia coli in the human urinary tract. Mol Microbiol. 2000;38(3):456-64. PubMed PMID: 11069670.

52. Cooke EM, Ewins SP. Properties of strains of Escherichia coli isolated from a variety of sources. J Med Microbiol. 1975;8(1):107-11. doi: 10.1099/00222615-8-1-107. PubMed PMID: 1092859.

53. Wiles TJ, Mulvey MA. The RTX pore-forming toxin alpha-hemolysin of uropathogenic Escherichia coli: progress and perspectives. Future Microbiol. 2013;8(1):73-84. doi: 10.2217/fmb.12.131. PubMed PMID: 23252494; PubMed Central PMCID: PMCPMC3570152.

54. van den Bosch JF, Postma P, de Graaff J, MacLaren DM. Haemolysis by urinary Escherichia coli and virulence in mice. J Med Microbiol. 1981;14(3):321-31. doi: 10.1099/00222615-14-3-321. PubMed PMID: 7021838.

55. Welch RA, Dellinger EP, Minshew B, Falkow S. Haemolysin contributes to virulence of extra-intestinal E. coli infections. Nature. 1981;294(5842):665-7. PubMed PMID: 7031483.

56. Wiles TJ, Bower JM, Redd MJ, Mulvey MA. Use of zebrafish to probe the divergent virulence potentials and toxin requirements of extraintestinal pathogenic Escherichia coli. PLoS Pathog. 2009;5(12):e1000697. doi: 10.1371/journal.ppat.1000697. PubMed PMID: 20019794; PubMed Central PMCID: PMC2785880.

57. Bucker R, Schulz E, Gunzel D, Bojarski C, Lee IF, John LJ, et al. alpha-Haemolysin of Escherichia coli in IBD: a potentiator of inflammatory activity in the colon. Gut. 2014;63(12):1893-901. doi: 10.1136/gutjnl-2013-306099. PubMed PMID: 24534723.

58. Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, Sabo A, et al. Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a

comparative genomics approach. Proc Natl Acad Sci U S A. 2006;103(15):5977-82. doi: 10.1073/pnas.0600938103. PubMed PMID: 16585510; PubMed Central PMCID: PMCPMC1424661.

59. Wijetunge DS, Karunathilake KH, Chaudhari A, Katani R, Dudley EG, Kapur V, et al. Complete nucleotide sequence of pRS218, a large virulence plasmid, that augments pathogenic potential of meningitis-associated Escherichia coli strain RS218. BMC Microbiol. 2014;14:203. doi: 10.1186/s12866-014-0203-9. PubMed PMID: 25164788; PubMed Central PMCID: PMCPMC4155114.

60. Cusumano CK, Hung CS, Chen SL, Hultgren SJ. Virulence plasmid harbored by uropathogenic Escherichia coli functions in acute stages of pathogenesis. Infect Immun. 2010;78(4):1457-67. doi: 10.1128/IAI.01260-09. PubMed PMID: 20123719; PubMed Central PMCID: PMCPMC2849428.

61. Lloyd AL, Rasko DA, Mobley HL. Defining genomic islands and uropathogen-specific genes in uropathogenic Escherichia coli. J Bacteriol. 2007;189(9):3532-46. doi: 10.1128/JB.01744-06. PubMed PMID: 17351047; PubMed Central PMCID: PMCPMC1855899.

62. Ritter A, Gally DL, Olsen PB, Dobrindt U, Friedrich A, Klemm P, et al. The Pai-associated leuX specific tRNA5(Leu) affects type 1 fimbriation in pathogenic Escherichia coli by control of FimB recombinase expression. Molecular microbiology. 1997;25(5):871-82. Epub 1997/11/19. PubMed PMID: 9364913.

63. Levin BR. The evolution and maintenance of virulence in microparasites. Emerg Infect Dis. 1996;2(2):93-102. doi: 10.3201/eid0202.960203. PubMed PMID: 8903208; PubMed Central PMCID: PMCPMC2639826.

64. Wold AE, Caugant DA, Lidin-Janson G, de Man P, Svanborg C. Resident colonic Escherichia coli strains frequently display uropathogenic characteristics. J Infect Dis. 1992;165(1):46-52. PubMed PMID: 1727897.

65. Diard M, Garry L, Selva M, Mosser T, Denamur E, Matic I. Pathogenicity-associated islands in extraintestinal pathogenic Escherichia coli are fitness elements involved in intestinal colonization. J Bacteriol. 2010;192(19):4885-93. doi: 10.1128/JB.00804-10. PubMed PMID: 20656906; PubMed Central PMCID: PMCPMC2944530.

66. Rodriguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, et al. The composition of the gut microbiota throughout life, with an emphasis on early life. Microb Ecol Health Dis. 2015;26:26050. doi: 10.3402/mehd.v26.26050. PubMed PMID: 25651996; PubMed Central PMCID: PMCPMC4315782.

67. Ogilvie LA, Jones BV. The human gut virome: a multifaceted majority. Front Microbiol. 2015;6:918. doi: 10.3389/fmicb.2015.00918. PubMed PMID: 26441861; PubMed Central PMCID: PMCPMC4566309.

68. Suhr MJ, Hallen-Adams HE. The human gut mycobiome: pitfalls and potentials--a mycologist's perspective. Mycologia. 2015;107(6):1057-73. doi: 10.3852/15-147. PubMed PMID: 26354806.

CHAPTER 2

TYPE I AND P PILI OF UROPATHOGENIC ESCHERICHIA COLI

Reprinted with permission from CAB International

Russell, C.W., Mulvey, M.A. (2014). Type 1 and P pili of uropathogenic *Escherichia coli*. In: Barocchi, M.A., Telford, J.L., editors. Bacterial pili: structure, synthesis and role in disease. CAB International, Wallingford, UK.

4 Type 1 and P Pili of Uropathogenic *Escherichia coli*

Colin Russell and Matthew Mulvey
Division of Microbiology and Immunology, University of Utah, Salt Lake City, USA

4.1 Introduction

Escherichia coli express a variety of proteinaceous structures known as pili or fimbriae that extend as hair-like fibres from the outer bacterial membrane. An adhesin protein located at the distal tip of each pilus fibre enables the bacteria to attach to various surfaces, facilitating colonization of environmental and host niches. In the case of uropathogenic *E. coli* (UPEC), type 1 pili (T1P) and P pili enable bacteria to bind to host epithelial cells within the urinary tract, which can lead to a urinary tract infection (UTI).

Several species of pathogenic bacteria can colonize and cause disease within the human urinary tract. However, UPEC are responsible for the overwhelming majority of UTIs (Ronald, 2002; Foxman, 2010). In general, UPEC colonize the urinary tract via an ascending route, moving from the urethra meatus through the urethra and into the bladder where they can cause cystitis (bladder infection) (Hannan *et al.*, 2012). UPEC can then climb the ureters and initiate pyelonephritis (kidney infection). Cystitis and pyelonephritis patients typically suffer from painful and frequent urination, while pyelonephritis patients can also experience more severe symptoms such as fever (Hooton, 2012). UTIs are one of the most common infections seen in the outpatient setting, accounting for well over 8 million doctor visits annually in the USA alone (Schappert and Rechtsteiner, 2011). Fifty percent of women will have a UTI by age 32 (Foxman and Brown, 2003), and 44% of patients will have another symptomatic episode within 1 year after initial infection (Ikaheimo *et al.*, 1996). The majority of UTI patients suffer from cystitis, whereas only 4% of patients develop pyelonephritis (Ikaheimo *et al.*, 1996). A small fraction of pyelonephritis infections will progress to urosepsis, in which UPEC disseminate systemically (Jolley *et al.*, 2012).

Colonization of the urinary tract by UPEC is expedited by T1P and P pili. UPEC attach to the bladder epithelium (or urothelium) via T1P, where they can subsequently trigger their internalization (Mulvey *et al.*, 1998; Martinez *et al.*, 2000). Within the cytosol of bladder superficial cells – which are binucleate and relatively large – the bacteria form T1P-dependent biofilm-like aggregates referred to as intracellular bacterial communities (IBC) (Eto *et al.*, 2006; Wright *et al.*, 2007; Jorgensen and Seed, 2012). As UPEC replicate within the host cell, individual bacterial cells can detach from the IBC, exit the host cell, and go on to bind, invade and colonize neighbouring superficial or underlying cells (Fig. 4.1) (Mulvey *et al.*, 2001; Justice *et al.*, 2004). In the case of

 Bacterial Pili: Structure, Synthesis and Role in Disease
(eds M.A. Barocchi and J.L. Telford)

pyelonephritis, P pili are important mediators of UPEC attachment to the kidney urothelium (O'Hanley *et al.*, 1985). In addition to their adhesive properties, T1P and P pili also trigger host responses that result in the release of cytokines and the influx of neutrophils (Fischer *et al.*, 2006; Godaly *et al.*, 2007).

4.2 Genetics and Structure of Type 1 and P Pili

T1P and P pili are encoded by the *fim* and *pap* loci, respectively (Fig. 4.2A). Each set of loci encodes an adhesin, structural and assembly proteins, and regulators of pilus expression. *fimH* and *papG* encode the adhesins of T1P and P pili, respectively, and are essential for pili function (Roberts *et al.*, 1994; Connell *et al.*, 1996). The structures of both pili are very similar (Fig. 4.2B), consisting of two parts: a helical rod and an adhesive fibrillar tip (Kuehn *et al.*, 1992; Jones *et al.*, 1995). The rods, which are comprised of 1000 or more FimA or PapA pilin subunits, are about 7 nm in diameter and vary from a few fractions of a micron to more than 5 μm in length. Each rod is anchored to the outer membrane through its interaction with the usher, a trans-membrane pore-forming protein (Thanassi *et al.*, 2012). Distally attached to each rod is a thinner (~3 nm), much shorter tip fibrillum made up of two to three adaptor proteins and an adhesin. During pilus biogenesis the structural subunits are translocated from the

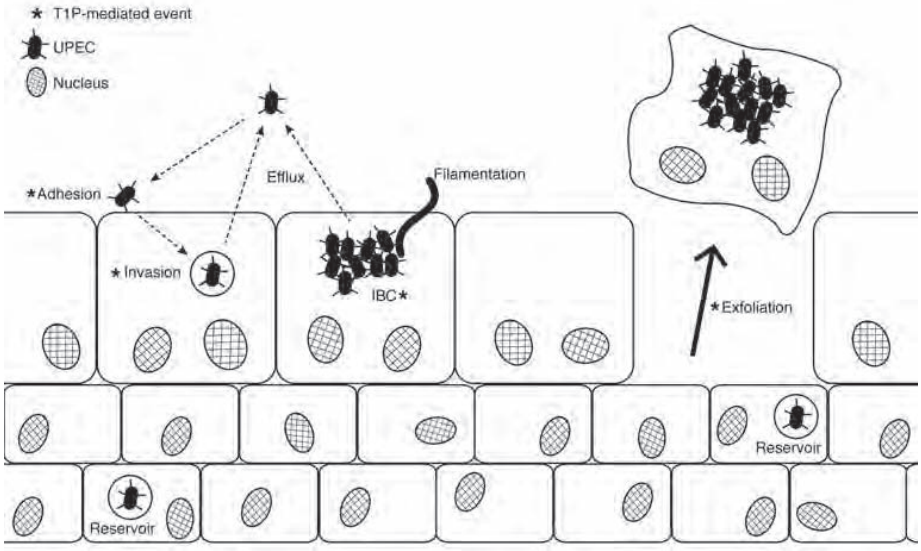


Fig. 4.1. Overview of bladder colonization by type 1-piliated UPEC. Pathogens enter the lumen of the bladder by an ascending route and subsequently bind and invade via T1P the large, binucleated superficial epithelial cells present on the apical surface of the urothelium. Some bacteria are quickly redirected back out into the bladder lumen, while others traffic to late endosome-like compartments where they can persist indefinitely in a quiescent state. Alternatively, UPEC can enter the host cytosol and rapidly multiply, forming IBCs. Infected superficial cells eventually die or exfoliate, taking with them large numbers of UPEC that can go on to infect other hosts or other sites within the urinary tract. Exfoliation, as well as the influx of immune cells like neutrophils, can compromise the barrier function of the urothelium. This allows UPEC to invade the smaller immature cells of the urothelium, where they can establish long-lived reservoirs that can persist in the presence of robust inflammatory responses and standard antibiotic treatments.

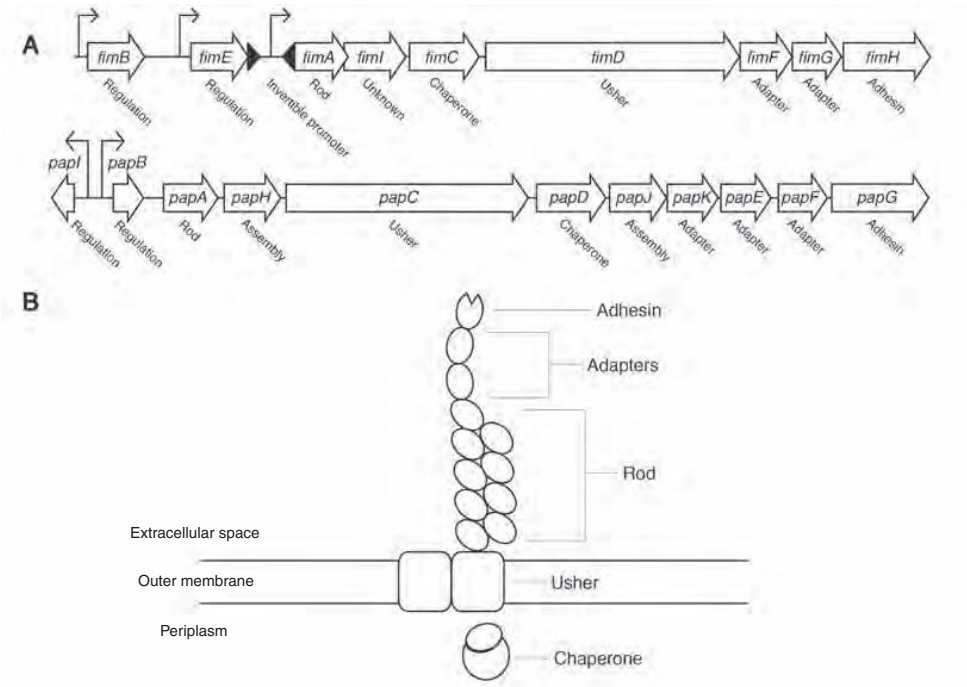


Fig. 4.2. (A) Organization of the *fim* and *pap* gene clusters, with (B) a basic model of pilus assembly by the chaperone–usher pathway.

bacterial cytosol into the periplasm. A chaperone then stabilizes and delivers the subunits to the usher protein in the outer membrane. Finally, the usher protein mediates addition of new subunits to the growing pilus (Allen *et al.*, 2012; Thanassi *et al.*, 2012).

4.3 Type 1 Pili

4.3.1 Adhesion

T1P-mediated adhesion of UPEC to superficial cells of the urothelium is one of the first steps during the infection process and can lead to host cell invasion, intracellular bacterial replication and persistence. FimH binds mannose on glycosylated host receptor proteins (Abraham *et al.*, 1985b; Krogfelt *et al.*, 1990). FimH binds mannose through a binding pocket within an N-terminal lectin domain, which is joined by a linker to the C-terminal pilin domain, the portion of the protein that connects FimH to the pilus (Choudhury, 1999; Schembri *et al.*, 2000).

The receptors engaged during the course of a UTI are not yet completely detailed, although several cell-surface proteins that bind T1P have been identified *in vitro*. FimH can bind to the uroplakin protein UPIa (Zhou *et al.*, 2001), which – along with at least three other uroplakin proteins – is important for maintaining the permeability barrier of the urothelium (Wu *et al.*, 2009). Because of its abundance on the urothelial surface, UPIa may be the predominant receptor that is used by UPEC to initiate infection. High-resolution electron microscopy showing T1P engaging the uroplakin-covered urothelium within the bladder of a mouse infected with UPEC yielded indirect *in vivo* evidence for T1P-

uroplakin interactions (Mulvey *et al.*, 1998). Cryo-electron microscopy demonstrated that FimH interactions with UP1a could trigger conformational changes within uroplakin complexes, possibly transducing signals into host bladder cells (Wang *et al.*, 2009).

Other proteins bound by T1P include the α3β1 integrin heterodimer, β2 integrin (CD18), CD48, the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family of proteins, glycoprotein 2 (GP2), Toll-like receptor 4 (TLR4), and the extracellular proteins fibronectin, collagen and laminin (Sauter *et al.*, 1991, 1993; Gbarah *et al.*, 1991; Sokurenko *et al.*, 1992; Kukkonen *et al.*, 1993; Baorto *et al.*, 1997; Pouttu *et al.*, 1999; Eto *et al.*, 2007; Mossman *et al.*, 2008; Carvalho *et al.*, 2009; Hase *et al.*, 2009; Yu and Lowe, 2009). These receptors exhibit distinct expression patterns and localization. α3β1 is located at cellular junctions and elsewhere throughout all layers of the urothelium (Southgate *et al.*, 1995). CEACAMs are also found on the urothelial surface (Kuespert *et al.*, 2006), and CEACAMs, β2 integrin and CD48 are present on the surface of immune cells such as neutrophils and macrophages (Gbarah *et al.*, 1991; Sauter *et al.*, 1991; Baorto *et al.*, 1997). Despite *in vitro* evidence for T1P interaction with these receptors, corroboration by *in vivo* experiments is still lacking.

Regardless of what host receptors T1P exploit, the interaction must be strong enough so as to withstand the shear forces produced by the bulk flow of urine during micturition. Interestingly, the receptor binding affinity of FimH rises as flow increases (Thomas *et al.*, 2002, 2004). Under low flow conditions the pilin domain of FimH inhibits the lectin domain in such a way that the mannose binding pocket is in a low affinity state (Le Trong *et al.*, 2010). As flow increases, the linker between the two domains stretches, releasing the inhibition on the lectin domain, allowing the binding pocket to adopt a high affinity conformation. Thus, as flow increases FimH binds more tightly to its receptor, ensuring that the bacteria are not dispelled into the urine.

Although *fimH* displays a high degree of conservation among various *E. coli* isolates (Abraham *et al.*, 1988; Weissman *et al.*, 2006), different *fimH* alleles do exist that vary in their abilities to bind mannose (Sokurenko *et al.*, 1995). This diversity in mannose-binding ability is linked with sequence variations in the lectin domain and linker, as well as alterations in composition of the type 1 pilus rod (Schembri *et al.*, 2000; Hung *et al.*, 2002; Mulvey, 2002; Thomas *et al.*, 2002). FimH from most T1P$^+$ *E. coli* – including those from the gut – are capable of binding trimannose, but UPEC-associated *fimH* alleles often have a higher binding affinity for monomannose residues (Sokurenko *et al.*, 1997, 1998). Phylogenetic evidence indicates that the *fimH* alleles carried by UPEC isolates evolved under positive selection from *fimH* alleles encoded by commensal strains within the intestinal microbiota (Sokurenko *et al.*, 1998; Weissman *et al.*, 2006). UPEC are believed to persist within the gut without eliciting any overt pathology, and many UPEC isolates found in urine closely match strains from the patient's colon (Russo *et al.*, 1995). Together, these data suggest a model in which faecal-borne UPEC strains that make their way into the urinary tract experience pressures that select *fimH* alleles that can mediate higher affinity interactions with receptors containing accessible monomannose residues. Since FimH makes up only a small portion of the gross T1P structure, it is theorized that *fimH* alleles in UPEC strains are not modified under selective pressure to evade antigenic recognition by the immune system (Weissman *et al.*, 2006). Rather, the composition and abundance of specific N-linked glycan receptors in the urinary tract probably help drive the selection process within individual hosts (Taganna *et al.*, 2011). Overall, these sorts of observations suggest that UPEC express *fimH* alleles that are uniquely evolved for life in the urinary tract.

4.3.2 Exfoliation

The presence of bacterial antigens within the bladder can stimulate the detachment of superficial cells from the urothelium (Mulvey *et al.*, 1998). This exfoliation process occurs via an apoptosis-like mechanism and provides the host a possible defence mechanism

to dispose of infected host cells via the urine (Aronson *et al.*, 1988; Mulvey *et al.*, 1998). However, exfoliation temporarily exposes the deeper portions of the urothelium. This gives the bacteria the opportunity to bind and invade the underlying immature intermediate and basal epithelial cells of the urothelium where they may be able to evade the immune system and establish long-lived reservoirs (see below) (Mulvey *et al.*, 2001).

In mouse models of UTI, FimH is necessary to cause exfoliation, and the expression of FimH by non-pathogenic laboratory *E. coli* strains allows them to induce exfoliation *in vivo*, similar to UPEC (Mulvey *et al.*, 1998). Binding of FimH to bladder epithelial cells can inhibit the anti-apoptotic function of the host transcription factor NF-κB, contributing to the induction of apoptotic pathways (Klumpp *et al.*, 2001, 2006). Binding of FimH to UPIa may stimulate the exfoliation process by triggering pro-apoptotic signalling cascades via effects on the cytoplasmic tail of neighbouring UPIIIa proteins (Thumbikat *et al.*, 2009). Secreted bacterial toxins like α-hemolysin and cytotoxin necrotizing factor 1 can also promote the exfoliation process, acting synergistically with type 1 pili (Mills *et al.*, 2000; Wiles *et al.*, 2008; Dhakal and Mulvey, 2012). While UPEC and laboratory *E. coli* strains that lack FimH fail to induce appreciably high levels of exfoliation *in vivo*, these bacterial strains can effectively trigger exfoliation in *ex vivo* experiments using bladder tissue explants (unpublished observations). These results indicate that type 1 pili can accelerate bladder cell exfoliation, possibly by keeping bacteria in close contact with host cells against the flow of urine, but other as-yet-undefined factors and signalling events likely initiate the process.

4.3.3 Invasion

T1P-mediated adhesion of bacteria to urothelial cells can lead to internalization via an endocytic-like mechanism in which the host plasma membrane zippers around and envelopes bound bacteria (Mulvey *et al.*, 1998; Martinez *et al.*, 2000). Bacteria are then trafficked into acidic, late endosome-like compartments via a pathway that involves Rab27b, a regulator of vesicular traffic that has been implicated in the delivery of uroplakin complexes to the host cell surface (Chen *et al.*, 2003; Eto *et al.*, 2006, 2008; Mysorekar and Hultgren, 2006; Bishop *et al.*, 2007). Within immature bladder epithelial cells, UPEC typically remains bound in the endosomal compartments, which are often enmeshed within a matrix of actin filaments (Eto *et al.*, 2006). These bacteria are highly resistant to antibiotics, many of which cannot cross host membranes and are ineffective against non-replicating microbes (Blango and Mulvey, 2010). UPEC can persist in a quiescent state within endosomal compartments for many days to weeks, serving as potential reservoirs for recurrent, relapsing and chronic UTIs that afflict many individuals (Mulvey *et al.*, 1998, 2001; Hvidberg *et al.*, 2000; Kerrn *et al.*, 2005; Mysorekar and Hultgren, 2006; Blango and Mulvey, 2010).

The FimH adhesin within T1P stimulates bacterial internalization by causing host receptors and signalling complexes to cluster and activate at sites of attachment, but the specific mechanisms remain incompletely defined (Martinez *et al.*, 2000; Dhakal *et al.*, 2008). Receptors implicated in the invasion process include UP1a and α3β1 integrin, though other receptors may also be involved (Eto *et al.*, 2007; Wang *et al.*, 2009). For example, complement and the host receptor for opsonized bacteria, CD46, can synergize with T1P to enhance bacterial internalization (Li *et al.*, 2009). The recruitment of host receptors and subsequent activation of signalling cascades that lead to UPEC internalization is apparently modulated by cholesterol-rich membrane domains known as lipid rafts, which serve as staging sites for many signalling events in the host plasma membrane (Duncan *et al.*, 2004; Eto *et al.*, 2008). T1P-mediated entry into host cells also requires actin rearrangements and, not surprisingly, many different host regulators of actin dynamics (Martinez *et al.*, 2000; Martinez and Hultgren, 2002; Duncan *et al.*, 2004; Eto *et al.*, 2007; Visvikis *et al.*, 2011). These include Src and focal adhesin kinases, phosphoinositides 3-kinase, Rho-family

to dispose of infected host cells via the urine (Aronson *et al.*, 1988; Mulvey *et al.*, 1998). However, exfoliation temporarily exposes the deeper portions of the urothelium. This gives the bacteria the opportunity to bind and invade the underlying immature intermediate and basal epithelial cells of the urothelium where they may be able to evade the immune system and establish long-lived reservoirs (see below) (Mulvey *et al.*, 2001).

In mouse models of UTI, FimH is necessary to cause exfoliation, and the expression of FimH by non-pathogenic laboratory *E. coli* strains allows them to induce exfoliation *in vivo*, similar to UPEC (Mulvey *et al.*, 1998). Binding of FimH to bladder epithelial cells can inhibit the anti-apoptotic function of the host transcription factor NF-κB, contributing to the induction of apoptotic pathways (Klumpp *et al.*, 2001, 2006). Binding of FimH to UPIa may stimulate the exfoliation process by triggering pro-apoptotic signalling cascades via effects on the cytoplasmic tail of neighbouring UPIIIa proteins (Thumbikat *et al.*, 2009). Secreted bacterial toxins like α-hemolysin and cytotoxin necrotizing factor 1 can also promote the exfoliation process, acting synergistically with type 1 pili (Mills *et al.*, 2000; Wiles *et al.*, 2008; Dhakal and Mulvey, 2012). While UPEC and laboratory *E. coli* strains that lack FimH fail to induce appreciably high levels of exfoliation *in vivo*, these bacterial strains can effectively trigger exfoliation in *ex vivo* experiments using bladder tissue explants (unpublished observations). These results indicate that type 1 pili can accelerate bladder cell exfoliation, possibly by keeping bacteria in close contact with host cells against the flow of urine, but other as-yet-undefined factors and signalling events likely initiate the process.

4.3.3 Invasion

T1P-mediated adhesion of bacteria to urothelial cells can lead to internalization via an endocytic-like mechanism in which the host plasma membrane zippers around and envelopes bound bacteria (Mulvey *et al.*, 1998; Martinez *et al.*, 2000). Bacteria are then trafficked into acidic, late endosome-like compartments via a pathway that involves Rab27b, a regulator of vesicular traffic that has been implicated in the delivery of uroplakin complexes to the host cell surface (Chen *et al.*, 2003; Eto *et al.*, 2006, 2008; Mysorekar and Hultgren, 2006; Bishop *et al.*, 2007). Within immature bladder epithelial cells, UPEC typically remains bound in the endosomal compartments, which are often enmeshed within a matrix of actin filaments (Eto *et al.*, 2006). These bacteria are highly resistant to antibiotics, many of which cannot cross host membranes and are ineffective against non-replicating microbes (Blango and Mulvey, 2010). UPEC can persist in a quiescent state within endosomal compartments for many days to weeks, serving as potential reservoirs for recurrent, relapsing and chronic UTIs that afflict many individuals (Mulvey *et al.*, 1998, 2001; Hvidberg *et al.*, 2000; Kerrn *et al.*, 2005; Mysorekar and Hultgren, 2006; Blango and Mulvey, 2010).

The FimH adhesin within T1P stimulates bacterial internalization by causing host receptors and signalling complexes to cluster and activate at sites of attachment, but the specific mechanisms remain incompletely defined (Martinez *et al.*, 2000; Dhakal *et al.*, 2008). Receptors implicated in the invasion process include UP1a and α3β1 integrin, though other receptors may also be involved (Eto *et al.*, 2007; Wang *et al.*, 2009). For example, complement and the host receptor for opsonized bacteria, CD46, can synergize with T1P to enhance bacterial internalization (Li *et al.*, 2009). The recruitment of host receptors and subsequent activation of signalling cascades that lead to UPEC internalization is apparently modulated by cholesterol-rich membrane domains known as lipid rafts, which serve as staging sites for many signalling events in the host plasma membrane (Duncan *et al.*, 2004; Eto *et al.*, 2008). T1P-mediated entry into host cells also requires actin rearrangements and, not surprisingly, many different host regulators of actin dynamics (Martinez *et al.*, 2000; Martinez and Hultgren, 2002; Duncan *et al.*, 2004; Eto *et al.*, 2007; Visvikis *et al.*, 2011). These include Src and focal adhesin kinases, phosphoinositides 3-kinase, Rho-family

4.3.5 Regulation of type 1 pili expression

T1P undergo phase variation, in which the expression of the *fim* operon can be turned on or off by an epigenetic mechanism (Abraham *et al.*, 1985a). *In vitro*, individual bacterial cells in a population will switch expression of T1P on and off at a low rate, resulting in a mixed population of $T1P^+$ and $T1P^-$ cells (Freitag *et al.*, 1985). Phase variation has also been observed in human infection (Kisielius *et al.*, 1989). The switch from one state to another is accomplished by a DNA recombination event in which the *fim* promoter is inverted by the recombinases FimB and FimE (Fig. 4.2) (Abraham *et al.*, 1985a). FimE primarily changes the orientation from on to off, and FimB changes the orientation equally from off to on and from on to off (Klemm, 1986; McClain *et al.*, 1991; Gally *et al.*, 1996).

The expression of T1P is controlled by several transcription factors (Table 4.1). T1P synthesis is modulated via transcriptional regulation of the recombinases, and by alteration of the conformation of DNA to facilitate recombination. The conditions that induce expression of T1P are not well elucidated, but studies of T1P regulators suggest that nutrient availability and environmental stresses play key roles.

T1P phase variation is also driven by crosstalk with P pili. The P pili regulator PapB dampens the expression of T1P by up-regulating expression of FimE and inhibiting the synthesis of FimB (Xia *et al.*, 2000; Holden *et al.*, 2006). Conversely, T1P expression is capable of inhibiting P pili expression (Snyder *et al.*, 2005). The complexity of these regulatory networks is further exasperated if one considers potential crosstalk with other systems, such as those that control bacterial motility and chemotaxis (Simms and Mobley, 2008; Cooper *et al.*, 2012).

4.3.6 Relevance of type 1 pili in pathogenesis

There is overwhelming evidence that T1P are important in animal models of UTI. UPEC strains that express T1P are able to infect mice better than non-piliated strains (Hultgren *et al.*, 1985; Schaeffer *et al.*, 1987). In competitive assays in which a specific mutant bacterial strain is mixed one-to-one with the wild-type parent strain and injected into mice, Keith *et al.* (1986) found that mutants lacking functional T1P are outcompeted by wild-type bacteria within the bladder. In non-competitive assays, *fimH* null mutants have reduced survival within the mouse urinary tract and are significantly less immunogenic (Connell *et al.*, 1996). Blocking T1P attachment to the murine urothelium using soluble FimH receptor analogues like α-D-mannosides inhibits UPEC colonization (Klein *et al.*, 2010; Cusumano *et al.*, 2011; Jiang *et al.*, 2012), as does passive immunization with antibodies directed against T1P (Abraham *et al.*, 1985). Immunization of animals with purified pili or, even better, the purified FimH adhesin or antigenically modified variants of FimH can provide significant protection against experimental UTI (Silverblatt and Cohen, 1979; Silverblatt *et al.*, 1982; Langermann, 1997; Poggio *et al.*, 2006; Karam *et al.*, 2012). These observations are underscored by results from an unbiased genetic screen in mice that showed T1P to be important for UPEC survival within the urinary tract (Bahrani-Mougeot *et al.*, 2002), and microarray analysis of bacteria taken straight from the urine of infected mice indicating that T1P are upregulated during UTI (Snyder *et al.*, 2004).

Evidence that T1P are crucial for UPEC colonization of the urinary tract in humans is less straightforward. Despite the fact that most UPEC isolates encode T1P (Norinder *et al.*, 2012), several studies have found that T1P are often absent on the surface of *E. coli* strains present in urine samples collected from patients with UTI (Ofek *et al.*, 1981; Pere *et al.*, 1987; Lim *et al.*, 1998; Hagan *et al.*, 2010). Such observations, coupled with the fact that T1P are also encoded by most commensal non-pathogenic *E. coli* isolates within the gut (Hagberg *et al.*, 1983), brings into question the role of T1P as virulence factors within the human host. The existence of a group of *E. coli* strains that cause asymptomatic bacteriuria (ABU) further questions the utility of T1P as mediators of bacterial colonization in the human urinary tract. ABU isolates can efficiently colonize humans, growing to very

Table 4.1 Regulators of T1P expression.

Regulator	Description	Relevant environmental signals	T1P +/–	Mechanism of T1P regulation	Reference
RpoS	Regulates genes important for response to stress and stationary phase	Expression of *fim* is inhibited during stationary phase	–	Inhibits transcription of *fimB*, inhibiting T1P expression	Dove *et al.* (1997)
LrhA	Inhibition of flagella, motility, chemotaxis	Reciprocal regulation of T1P and flagella. Inhibits adherence when motility genes activated, and vice-versa	–	Activates transcription of *fimE*	Blumer *et al.* (2005), Simms and Mobley (2008)
ppGpp	Regulates the stringent response (during nutrient deprivation or environmental stress)	Amino acid shortage, etc.	+	Interacts with RNA polymerase to activate transcription of *fimB*	Blomfield *et al.* (1993), Aberg *et al.* (2008)
DksA	Regulates the stringent response (during nutrient deprivation or environmental stress)	Augments ppGpp effects	+	Interacts with RNA polymerase to activate transcription of *fimB*	Aberg *et al.* (2008)
leuX	tRNA that recognizes rare codons for leucine to facilitate translation		+	Facilitates the translation of *fimB* by recognizing rare leucine-encoding codons in *fimB* mRNA	Newman *et al.* (1994)
Lrp	Regulates expression of genes important for survival during nutrient deprivation; plays a role in DNA packaging	Recombination is stimulated by presence of aliphatic amino acids	+/–	Facilitates recombination by bending DNA. It is biased towards the ON phase	Gally *et al.* (1993)
IHF	Integration host factor, modulates transcription of genes by bending DNA		+	Alters DNA conformation to facilitate recombination towards the ON phase	Corcoran and Dorman (2009)
H-NS	Supercoils DNA to silence genes with high A+T content		–	Supercoils DNA, facilitates recombination towards the OFF phase	O'Gara and Dorman (2000)
CRP	Regulates genes for catabolism of secondary carbon sources	Growth in presence of glucose increases T1P	–	Stimulates DNA gyrase activity which supercoils DNA and inhibits transcription	Muller *et al.* (2009)

high numbers within the lumen of the bladder without causing any overt inflammatory responses and without binding to or invading urothelial cells (Hull *et al.*, 1999; Bergsten *et al.*, 2005; Klemm *et al.*, 2007). These strains typically lack functional T1P and P pili, as well as many other factors that are often associated with UPEC isolates (Klemm *et al.*, 2006). ABU strains can outcompete more pathogenic UPEC isolates, and are being developed as therapeutic tools to interfere with UPEC colonization of the host (Falagas *et al.*, 2008). The engineered expression of recombinant T1P by the prototypical ABU strain 83972 causes increased host inflammatory responses in a mouse UTI model, but this effect was not observed in human volunteers (Bergsten *et al.*, 2007). Furthermore, an older study concluded that the 83972 strain expressing T1P was rapidly outcompeted by the non-adherent wild-type 83972 strain in the human urinary tract (Andersson *et al.*, 1991). While informative, these results are difficult to interpret because the patient volunteers had complicated histories of recurrent or chronic UTIs that were unresponsive to antibiotic therapy.

Overall, these results indicate that T1P are not an absolute requirement for bacterial colonization of the human urinary tract. For example, ABU isolates can apparently employ alternate T1P-independent strategies to persist within the host, probably taking advantage of more robust metabolic pathways present within their streamlined genomes to grow more efficiently within the extracellular urine-saturated milieu of the bladder. In the case of *bona fide* UPEC isolates, the necessity of T1P remains debatable. While UPEC strains present in the urine of patients with UTI do not always have T1P present, the expression of these fibres can almost always be induced when the strains are grown in broth culture (Ofek *et al.*, 1981; Pere *et al.*, 1987; Lim *et al.*, 1998; Hagan *et al.*, 2010). It can therefore be argued that the free bacteria present in the urine simply have their T1P in the off phase, while those in the on phase remain associated with the urothelium and allow UPEC to persist within the urinary tract and ultimately cause disease (Hultgren *et al.*, 1985; Lim *et al.*, 1998). Indeed, evidence that T1P-mediated adherence, host cell invasion and IBC development are critical to the progression of UTIs in humans is mounting, as exemplified by data showing the remnants of IBCs within urothelial cells that are shed into the urine of women suffering from UTI (Rosen *et al.*, 2007).

4.4 P Pili

4.4.1 Adhesion

P pili play a part in colonization of the urinary tract by mediating adhesion to the Galα1-4Gal disaccharide moiety found on globo-series glycosphingolipids located on the surface of host cell membranes (Leffler and Svanborg-Eden, 1981; O'Hanley *et al.*, 1985; Wold *et al.*, 1988). Because there is a strong correlation between P pili and pyelonephritis, P pili have mostly been studied in the context of kidney infection (Kallenius *et al.*, 1981; Plos *et al.*, 1990). However, Galα1-4Gal glycolipids are found throughout the entire urothelium (O'Hanley *et al.*, 1985), suggesting that P pili may play a role in cystitis in some instances. The binding of P pili to glycosphingolipids in the urothelium is currently the only known function of P pili. They are dispensable for other events such as host cell invasion (Martinez *et al.*, 2000).

Like T1P, P pili may enable UPEC to better withstand the shear flow of urine. P pili can unwind *in vitro* to five times their length in a reversible manner (Gong and Makowski, 1992; Bullitt and Makowski, 1995; Thanassi *et al.*, 1998; Fallman *et al.*, 2005). It is hypothesized that this unwinding gives the pilus more flexibility during urine flow, preventing the pilus from breaking or detaching from the urothelium.

The adhesive tip subunit of P pili, encoded by *papG*, is highly polymorphic. Distinct alleles of *papG* correlate with UPEC strains isolated from different host species, as well as specific host-associated niches. These alleles have differing affinities for Galα1-4Gal, depending on the topography of the glycolipid receptors present (Stromberg *et al.*, 1990). Class II *papG* alleles are found in human pyelonephritis and urosepsis isolates

(Otto *et al.*, 1993; Johanson *et al.*, 1993; Johnson, 1998; Johnson *et al.*, 1998a), while class III alleles are encoded by human cystitis isolates as well as canine UTI strains (Johanson *et al.*, 1993; Johnson *et al.*, 1998b, 2000). Other *papG* allele classes are less common and little is known about their binding propensities.

4.4.2 Regulation of P pili expression

Similar to T1P and most other fimbrial adhesins, P pili are subject to phase variation, switching between an on and off state (Blyn *et al.*, 1989; Kisielius *et al.*, 1989). The frequency of phase switching can vary greatly dependent upon strain background and culture conditions (Holden *et al.*, 2007). Unlike T1P, phase variation in P pili is driven by differential methylation of the *pap* promoter and not by DNA recombination (Blyn *et al.*, 1990). Methylation occurs at two GATC sequences within the promoter. GATC sites are found throughout the entire genome and are typically recognized and methylated by the enzyme DNA adenine methyltransferase (Dam). When the site proximal to the *papBA* operon ($GATC^{prox}$) is methylated, *pap* genes are transcribed and P pili are produced (Fig. 4.3). Conversely, when $GATC^{prox}$ is unmethylated and the distal GATC site ($GATC^{dist}$) is methylated, the main *pap* promoter P_{BA} is not activated and P pili are not made.

Binding of the Leucine-responsive protein (Lrp) modulates the differential methylation of the *pap* promoter in an autoregulatory fashion by competing with Dam for binding sites within the *pap* (P_{BA}) promoter (Braaten *et al.*, 1992; Kawamura *et al.*, 2011). There are six binding sites for Lrp within the *pap* promoter (Fig. 4.3), two of which overlap the $GATC^{prox}$ and $GATC^{dist}$ methylation sites. When $GATC^{prox}$ is methylated, Lrp binds more readily to $GATC^{dist}$ where it can stimulate expression of PapI (Goransson *et al.*, 1989). PapI promotes interactions between Lrp and distal Lrp binding sites within the promoter, allowing Dam and RNA polymerase access to proximal

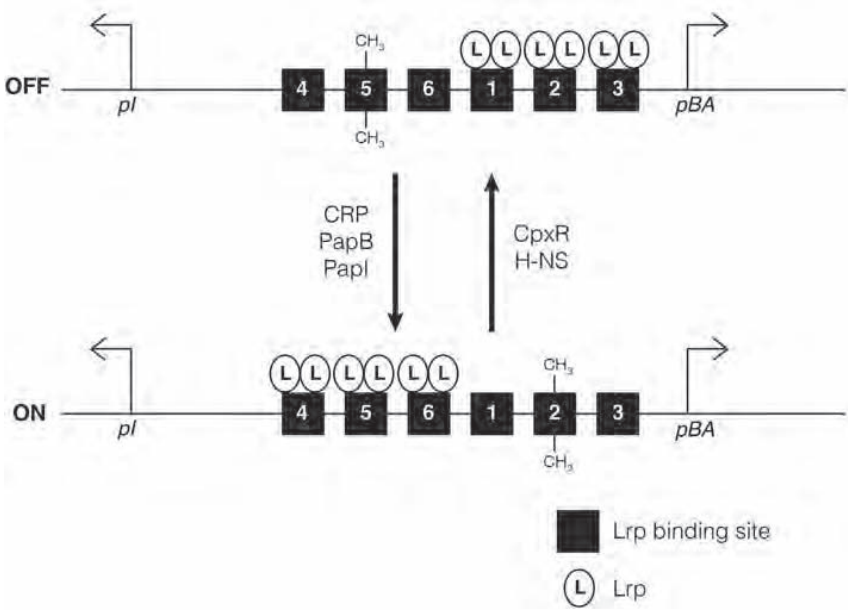


Fig. 4.3. Model of phase switching of *pap* genes by multiple regulators. Methylation (CH_3) sites and promoters for *papI* and *papBA* are indicated.

sites upstream of *papBA* and thereby activating P pili expression (Nou *et al.*, 1993; Kaltenbach *et al.*, 1995; Kawamura *et al.*, 2011). Increased expression of PapB can result in re-occupation of the proximal, higher affinity Lrp binding sites, resulting in the inhibition of P pili expression. Although Lrp typically activates genes important for amino acid metabolism and nutrient transport (Newman and Lin, 1995), the activity of Lrp within the *pap* gene cluster is not responsive to leucine (Braaten *et al.*, 1992).

Transcription of *papBA* is inhibited by the response regulator CpxR, part of the Cpx two-component system that senses and responds to membrane stress and unfolded periplasmic proteins (Hung *et al.*, 2001; Hernday *et al.*, 2004). When misfolded pilin subunits are sensed, CpxR downregulates the expression of P pili by blocking Lrp-binding (Hernday *et al.*, 2004). The *papBA* operon is also repressed by the histone-like nucleoid-structuring protein H-NS (Goransson *et al.*, 1990; White-Ziegler *et al.*, 1998), a global regulator that can preferentially silence the transcription of A+T-rich DNA sequences by inducing DNA supercoiling (Fang and Rimsky, 2008). Repression by H-NS is relieved by PapB and PapB-mediated effects on PapI expression, as well as Catabolite Responsive Protein (CRP) (Baga *et al.*, 1985; Forsman *et al.*, 1989, 1992; Goransson *et al.*, 1989). CRP enables UPEC and other bacteria to respond to changing carbon sources and may act to enhance P pili expression when glucose availability is low (Baga *et al.*, 1985; Goransson *et al.*, 1989; Forsman *et al.*, 1992; Weyand *et al.*, 2001).

DNA replication itself can also affect phase variable expression of the *pap* operon. After DNA replication, Dam methylates the new, unmethylated DNA strand at most of the GATC sites in the genome. However, GATC sites within the *pap* promoter are not optimal targets for Dam (Peterson and Reich, 2006). Slower methylation of sites within the *pap* promoter enables Lrp to competitively bind to the GATC sequences, blocking their methylation by Dam and repressing P pili expression (Hernday *et al.*, 2003; Peterson and Reich, 2008). These regulatory effects are dependent upon the concentrations of Lrp, and are likely complicated further by myriad regulatory factors that come into play during the course of a UTI.

4.4.3 Relevance of P pili in pathogenesis

The importance of P pili in UTI has been shown in rodents, monkeys and humans. In mice, the delivery of P pili-specific antibodies provides passive protection against kidney colonization by UPEC (O'Hanley *et al.*, 1985). UPEC strains, and even faecal isolates, that express recombinant P pili can outcompete P pili-negative counterparts within the kidneys of mice (Hagberg *et al.*, 1983), but in other work P pili expression in mice was inconsequential to the ability of a prototypic UPEC isolate to cause pyelonephritis (Mobley *et al.*, 1993). In monkeys, immunization with purified P pili can interfere with the ability of UPEC to cause pyelonephritis (Roberts *et al.*, 1984), and in at least one instance disruption of the *papG* allele was shown to attenuate UPEC colonization of monkey kidneys (Roberts *et al.*, 1994). Corroborating these primate studies is work showing that human pyelonephritic strains are more likely to have the *pap* genes and express P pili than cystitis, ABU and faecal isolates (Kallenius *et al.*, 1981; Plos *et al.*, 1990). In some studies with human volunteers, transformation of the ABU strain 83972 with *pap* genes enhanced bacterial growth, survival and inflammatory effects within the urinary tract (Wullt *et al.*, 2000, 2001; Wullt, 2003), but no effects were seen in a similar, older study (Andersson *et al.*, 1991). These varying results may be attributable to differences in the expression constructs used, the specific *pap* alleles examined and the medical histories of the patient volunteers.

In total, these results from animal and human studies indicate that P pili expression can impact the establishment and progression of kidney infections, but this may vary dependent on both strain and host background. The precise mechanisms by which P pili promote bacterial colonization of the kidneys is not entirely clear, though recent work employing intravital imaging of infected rat kidneys suggest that P pili may act in a synergistic fashion with T1P within

the proximal tubules of the kidneys (Melican *et al.*, 2011). Specifically, P pili seem to promote early colonization of the kidney tubules, allowing UPEC to resist flow generated by filtration processes within the kidneys. Later, T1P expression may help adherent bacteria within the tubules assemble into biofilm-like aggregates that are even more resistant to elimination.

4.5 Type 1 Pili, P Pili and the Immune Response

The immune response to UPEC is mostly mediated by innate defences, with T and B cells and other components of adaptive immunity playing a smaller and less well-studied role (Ragnarsdottir and Svanborg, 2012). The innate response involves activation of pattern recognition receptors such as TLR4 and subsequent upregulation of pro-inflammatory responses, including increased expression of cytokines like IL-6 and IL-8 (Frendeus *et al.*, 2001; Ragnarsdottir *et al.*, 2011). IL-6 stimulates fever, increases the production of C-reactive protein to activate complement and stimulates B cells to secrete mucosal IgG (Ragnarsdottir and Svanborg, 2012). IL-8 is a chemokine important for recruiting neutrophils to sites of infection where they can target and kill bacteria (Haraoka *et al.*, 1999). These inflammatory responses can be initiated by a panoply of pathogen-associated molecules, including lipopolysaccharide (LPS). LPS binds and activates TLR4 via association with the TLR4 co-receptor CD14 (Palsson-Mcdermott and O'Neill, 2004). Interestingly, human urothelial cells in both the kidneys and bladder lack membrane-bound CD14, but are still responsive to UPEC (Samuelsson *et al.*, 2004). *In vitro* studies suggest that FimH, by binding TLR4 directly, may inadvertently present LPS to TLR4 and thereby bypasses the need for CD14 (Hedlund *et al.*, 2001; Samuelsson *et al.*, 2004; Fischer *et al.*, 2006; Mossman *et al.*, 2008). P pili can also stimulate TLR4 signalling, though through a more circuitous pathway (Fischer *et al.*, 2007, 2010). Rather than interacting with TLR4 directly, the engagement of glycosphingolipid receptors by P pili leads to the release of ceramide from the host cell and this in turn binds and activates TLR4 even in the absence of LPS (Hedlund *et al.*, 1999; Fischer *et al.*, 2007, 2010). These processes may enable the urothelium to modulate its sensitivity to LPS, avoiding the activation of potentially damaging inflammatory responses unless truly virulent T1P- or P pili-positive pathogens are present (Samuelsson *et al.*, 2004).

Activation of TLR4 via T1P and P pili may also benefit UPEC by eliciting tissue damage that could facilitate bacterial dissemination within the host. UPEC strains display an ability to survive longer than non-pathogenic *E. coli* strains within macrophages, neutrophils and mast cells, and could potentially hijack these immune effector cells as an additional means to spread within the urinary tract (Baorto *et al.*, 1997; Shin, 2000; Nazareth *et al.*, 2007; Bokil *et al.*, 2011). The ability of some UPEC isolates to dampen inflammatory responses under some conditions may enable the pathogens to fine-tune the host environment over the course of an infection (Hunstad *et al.*, 2005; Billips *et al.*, 2007; Wiles *et al.*, 2008; Hilbert *et al.*, 2008; Loughman and Hunstad, 2011, 2012; Dhakal and Mulvey, 2012). This is, however, a potentially dangerous game for UPEC, as T1P interactions with receptors on neutrophils, macrophages and mast cells can stimulate oxidative burst and other antibacterial responses (Sauter *et al.*, 1991; Tewari *et al.*, 1993; Malaviya *et al.*, 1999; Ashkar *et al.*, 2008). These can include the release of antimicrobial peptides, which can work in conjunction with soluble proteins like Tamm-Horsfall Protein (THP) and secretory IgA (sIgA) that can bind FimH and block UPEC interactions with host cells (Wold *et al.*, 1990; Bates *et al.*, 2004; Mo *et al.*, 2004).

4.6 Conclusion

Data accumulated over the past few decades indicate that T1P and P pili are preeminent virulence and fitness factors utilized by many UPEC isolates to effectively colonize host tissues within the urinary tract. Ongoing work demonstrates that these adhesive

organelles can function in a complex array of biological activities, ranging from epithelial cell attachment and invasion to biofilm formation and the modulation of host inflammatory responses. The central roles that these pili have as mediators of UPEC pathogenesis within the urinary tract has made them prime targets for the development of vaccines and other antimicrobial therapeutics. Future studies will undoubtedly highlight new ways in which these organelles can affect the onset and progression of UTIs. It is likewise probable that the coming years will reveal effects that T1P and P pili have on the fitness of both pathogens and commensal strains at sites outside of the urinary tract.

References

Aberg, A., Shingler, V. and Balsalobre, C. (2008) Regulation of the *fimB* promoter: a case of differential regulation by ppGpp and DksA *in vivo*. *Molecular Microbiology* 67, 1223–1241.

Abraham, J.M., Freitag, C.S., Clements, J.R. and Eisenstein, B.I. (1985a) An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 82, 5724–5727.

Abraham, S.N., Babu, J.P., Giampapa, C.S., Hasty, D.L., Simpson, W.A. and Beachey, E.H. (1985b) Protection against *Escherichia coli*-induced urinary tract infections with hybridoma antibodies directed against type 1 fimbriae or complementary D-mannose receptors. *Infection and Immunity* 48, 625–628.

Abraham, S.N., Sun, D., Dale, J.B. and Beachey, E.H. (1988) Conservation of the D-mannose-adhesion protein among type 1 fimbriated members of the family Enterobacteriaceae. *Nature* 336, 682–684.

Allen, W.J., Phan, G. and Waksman, G. (2012) Pilus biogenesis at the outer membrane of Gram-negative bacterial pathogens. *Current Opinion in Structural Biology* 22, 500–506.

Anderson, G.G., Palermo, J.J., Schilling, J.D., Roth, R., Heuser, J. and Hultgren, S.J. (2003) Intracellular bacterial biofilm-like pods in urinary tract infections. *Science* 301, 105–107.

Andersson, P., Engberg, I., Lidin-Janson, G., Lincoln, K., Hull, R., Hull, S. and Svanborg, C. (1991) Persistence of *Escherichia coli* bacteriuria is not determined by bacterial adherence. *Infection and Immunity* 59, 2915–2921.

Aronson, M., Medalia, O., Amichay, D. and Nativ, O. (1988) Endotoxin-induced shedding of viable uroepithelial cells is an antimicrobial defense mechanism. *Infection and Immunity* 56, 1615–1617.

Ashkar, A.A., Mossman, K.L., Coombes, B.K., Gyles, C.L. and Mackenzie, R. (2008) FimH adhesin of type 1 fimbriae is a potent inducer of innate antimicrobial responses which requires TLR4 and type 1 interferon signalling. *PLoS Pathogens* 4, e1000233.

Baga, M., Goransson, M., Normark, S. and Uhlin, B.E. (1985) Transcriptional activation of a pap pilus virulence operon from uropathogenic *Escherichia coli*. *EMBO Journal* 4, 3887–3893.

Bahrani-Mougeot, F.K., Buckles, E.L., Lockatell, C.V., Hebel, J.R., Johnson, D.E., Tang, C.M. and Donnenberg, M.S. (2002) Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia coli* virulence determinants in the murine urinary tract. *Molecular Microbiology* 45, 1079–1093.

Baorto, D.M., Gao, Z., Malaviya, R., Dustin, M.L., Van Der Merwe, A., Lublin, D.M. and Abraham, S.N. (1997) Survival of FimH-expressing enterobacteria in macrophages relies on glycolipid traffic. *Nature* 389, 636–639.

Bates, J.M., Raffi, H.M., Prasadan, K., Mascarenhas, R., Laszik, Z., Maeda, N., Hultgren, S.J. and Kumar, S. (2004) Tamm-Horsfall protein knockout mice are more prone to urinary tract infection: rapid communication. *Kidney International* 65, 791–797.

Bergsten, G., Wullt, B. and Svanborg, C. (2005) *Escherichia coli*, fimbriae, bacterial persistence and host response induction in the human urinary tract. *International Journal of Medical Microbiology: IJMM* 295, 487–502.

Bergsten, G., Wullt, B., Schembri, M.A., Leijonhufvud, I. and Svanborg, C. (2007) Do type 1 fimbriae promote inflammation in the human urinary tract? *Cellular Microbiology* 9, 1766–1781.

Billips, B.K., Forrestal, S.G., Rycyk, M.T., Johnson, J.R., Klumpp, D.J. and Schaeffer, A.J. (2007) Modulation of host innate immune response in the bladder by uropathogenic *Escherichia coli*. *Infection and Immunity* 75, 5353–5360.

Bishop, B.L., Duncan, M.J., Song, J., Li, G., Zaas, D. and Abraham, S.N. (2007) Cyclic AMP-regulated exocytosis of *Escherichia coli* from infected bladder epithelial cells. *Nature Medicine* 13, 625–630.

Blango, M.G. and Mulvey, M.A. (2010) Persistence of uropathogenic *Escherichia coli* in the face of multiple antibiotics. *Antimicrobial Agents and Chemotherapy* 54, 1855–1863.

Blomfield, I.C., Calie, P.J., Eberhardt, K.J., McClain, M.S. and Eisenstein, B.I. (1993) Lrp stimulates phase variation of type 1 fimbriation in *Escherichia coli* K-12. *Journal of Bacteriology* 175, 27–36.

Blumer, C., Kleefeld, A., Lehnen, D., Heintz, M., Dobrindt, U., Nagy, G., Michaelis, K., Emody, L., Polen, T., Rachel, R., Wendisch, V.F. and Unden, G. (2005) Regulation of type 1 fimbriae synthesis and biofilm formation by the transcriptional regulator LrhA of *Escherichia coli*. *Microbiology* 151, 3287–3298.

Blyn, L.B., Braaten, B.A., White-Ziegler, C.A., Rolfson, D.H. and Low, D.A. (1989) Phase-variation of pyelonephritis-associated pili in *Escherichia coli*: evidence for transcriptional regulation. *EMBO Journal* 8, 613–620.

Blyn, L.B., Braaten, B.A. and Low, D.A. (1990) Regulation of pap pilin phase variation by a mechanism involving differential dam methylation states. *EMBO Journal* 9, 4045–4054.

Bokil, N.J., Totsika, M., Carey, A.J., Stacey, K.J., Hancock, V., Saunders, B.M., Ravasi, T., Ulett, G.C., Schembri, M.A. and Sweet, M.J. (2011) Intramacrophage survival of uropathogenic *Escherichia coli*: differences between diverse clinical isolates and between mouse and human macrophages. *Immunobiology* 216, 1164–1171.

Braaten, B.A., Platko, J.V., Van Der Woude, M.W., Simons, B.H., De Graaf, F.K., Calvo, J.M. and Low, D.A. (1992) Leucine-responsive regulatory protein controls the expression of both the pap and fan pili operons in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* 89, 4250–4254.

Bullitt, E. and Makowski, L. (1995) Structural polymorphism of bacterial adhesion pili. *Nature* 373, 164–167.

Carvalho, F.A., Barnich, N., Sivignon, A., Darcha, C., Chan, C.H., Stanners, C.P. and Darfeuille-Michaud, A. (2009) Crohn's disease adherent-invasive *Escherichia coli* colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM. *Journal of Experimental Medicine* 206, 2179–2189.

Chen, Y., Guo, X., Deng, F.M., Liang, F.X., Sun, W., Ren, M., Izumi, T., Sabatini, D.D., Sun, T.T. and Kreibich, G. (2003) Rab27b is associated with fusiform vesicles and may be involved in targeting uroplakins to urothelial apical membranes. *Proceedings of the National Academy of Sciences of the United States of America* 100, 14012–14017.

Choudhury, D. (1999) X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* 285, 1061–1066.

Connell, I., Agace, W., Klemm, P., Schembri, M., Marild, S. and Svanborg, C. (1996) Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proceedings of the National Academy of Sciences of the United States of America* 93, 9827–9832.

Cooper, L.A., Simmons, L.A. and Mobley, H.L. (2012) Involvement of mismatch repair in the reciprocal control of motility and adherence of uropathogenic *Escherichia coli*. *Infection and Immunity* 80, 1969–1979.

Corcoran, C.P. and Dorman, C.J. (2009) DNA relaxation-dependent phase biasing of the fim genetic switch in *Escherichia coli* depends on the interplay of H-NS, IHF and LRP. *Molecular Microbiology* 74, 1071–1082.

Cusumano, C.K., Pinkner, J.S., Han, Z., Greene, S.E., Ford, B.A., Crowley, J.R., Henderson, J.P., Janetka, J.W. and Hultgren, S.J. (2011) Treatment and prevention of urinary tract infection with orally active FimH inhibitors. *Science Translational Medicine* 3, 109ra115.

Dhakal, B.K. and Mulvey, M.A. (2009) Uropathogenic *Escherichia coli* invades host cells via an HDAC6-modulated microtubule-dependent pathway. *Journal of Biological Chemistry* 284, 446–454.

Dhakal, B.K. and Mulvey, M.A. (2012) The UPEC pore-forming toxin alpha-hemolysin triggers proteolysis of host proteins to disrupt cell adhesion, inflammatory, and survival pathways. *Cell Host & Microbe* 11, 58–69.

Dhakal, B.K., Kulesus, R.R. and Mulvey, M.A. (2008) Mechanisms and consequences of bladder cell invasion by uropathogenic *Escherichia coli*. *European Journal of Clinical Investigation* 38 Supplement 2, 2–11.

Dove, S.L., Smith, S.G. and Dorman, C.J. (1997) Control of *Escherichia coli* type 1 fimbrial gene expression in stationary phase: a negative role for RpoS. *Molecular & General Genetics* 254, 13–20.

Duncan, M.J., Li, G., Shin, J.S., Carson, J.L. and Abraham, S.N. (2004) Bacterial penetration of bladder epithelium through lipid rafts. *Journal of Biological Chemistry* 279, 18944–18951.

Eto, D.S., Sundsbak, J.L. and Mulvey, M.A. (2006) Actin-gated intracellular growth and resurgence of uropathogenic *Escherichia coli*. *Cellular Microbiology* 8, 704–717.

Eto, D.S., Jones, T.A., Sundsbak, J.L. and Mulvey, M.A. (2007) Integrin-mediated host cell invasion by type 1-piliated uropathogenic *Escherichia coli*. *PLoS Pathogens* 3, e100.

Eto, D.S., Gordon, H.B., Dhakal, B.K., Jones, T.A. and Mulvey, M.A. (2008) Clathrin, AP-2, and the NPXY-binding subset of alternate endocytic adaptors facilitate FimH-mediated bacterial invasion of host cells. *Cellular Microbiology* 10, 2553–2567.

Falagas, M.E., Rafailidis, P.I. and Makris, G.C. (2008) Bacterial interference for the prevention and treatment of infections. *International Journal of Antimicrobial Agents* 31, 518–522.

Fallman, E., Schedin, S., Jass, J., Uhlin, B.E. and Axner, O. (2005) The unfolding of the P pili quaternary structure by stretching is reversible, not plastic. *EMBO Reports* 6, 52–56.

Fang, F.C. and Rimsky, S. (2008) New insights into transcriptional regulation by H-NS. *Current Opinion in Microbiology* 11, 113–120.

Fischer, H., Yamamoto, M., Akira, S., Beutler, B. and Svanborg, C. (2006) Mechanism of pathogen-specific TLR4 activation in the mucosa: fimbriae, recognition receptors and adaptor protein selection. *European Journal of Immunology* 36, 267–277.

Fischer, H., Ellstrom, P., Ekstrom, K., Gustafsson, L., Gustafsson, M. and Svanborg, C. (2007) Ceramide as a TLR4 agonist; a putative signalling intermediate between sphingolipid receptors for microbial ligands and TLR4. *Cellular Microbiology* 9, 1239–1251.

Fischer, H., Lutay, N., Ragnarsdottir, B., Yadav, M., Jonsson, K., Urbano, A., Al Hadad, A., Ramisch, S., Storm, P., Dobrindt, U., Salvador, E., Karpman, D., Jodal, U. and Svanborg, C. (2010) Pathogen specific, IRF3-dependent signaling and innate resistance to human kidney infection. *PLoS Pathogens*, 6, e1001109.

Forsman, K., Goransson, M. and Uhlin, B.E. (1989) Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in *E. coli* pili biogenesis. *EMBO Journal* 8, 1271–1277.

Forsman, K., Sonden, B., Goransson, M. and Uhlin, B.E. (1992) Antirepression function in *Escherichia coli* for the cAMP-cAMP receptor protein transcriptional activator. *Proceedings of the National Academy of Sciences of the United States of America* 89, 9880–9884.

Foxman, B. (2010) The epidemiology of urinary tract infection. *Nature Reviews Urology* 7, 653–660.

Foxman, B. and Brown, P. (2003) Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. *Infectious Disease Clinics of North America* 17, 227–241.

Freitag, C.S., Abraham, J.M., Clements, J.R. and Eisenstein, B.I. (1985) Genetic analysis of the phase variation control of expression of type 1 fimbriae in *Escherichia coli*. *Journal of Bacteriology* 162, 668–675.

Frendeus, B., Wachtler, C., Hedlund, M., Fischer, H., Samuelsson, P., Svensson, M. and Svanborg, C. (2001) *Escherichia coli* P fimbriae utilize the Toll-like receptor 4 pathway for cell activation. *Molecular Microbiology* 40, 37–51.

Gally, D.L., Bogan, J.A., Eisenstein, B.I. and Blomfield, I.C. (1993) Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in *Escherichia coli* K-12: effects of temperature and media. *Journal of Bacteriology* 175, 6186–6193.

Gally, D.L., Leathart, J. and Blomfield, I.C. (1996) Interaction of FimB and FimE with the fim switch that controls the phase variation of type 1 fimbriae in *Escherichia coli* K-12. *Molecular Microbiology* 21, 725–738.

Gbarah, A., Gahmberg, C.G., Ofek, I., Jacobi, U. and Sharon, N. (1991) Identification of the leukocyte adhesion molecules CD11 and CD18 as receptors for type 1-fimbriated (mannose-specific) *Escherichia coli*. *Infection and Immunity* 59, 4524–4530.

Godaly, G., Otto, G., Burdick, M.D., Strieter, R.M. and Svanborg, C. (2007) Fimbrial lectins influence the chemokine repertoire in the urinary tract mucosa. *Kidney International* 71, 778–786.

Gong, M. and Makowski, L. (1992) Helical structure of P pili from *Escherichia coli*. Evidence from X-ray fiber diffraction and scanning transmission electron microscopy. *Journal of Molecular Biology* 228, 735–742.

Goransson, M., Forsman, P., Nilsson, P. and Uhlin, B.E. (1989) Upstream activating sequences that are shared by two divergently transcribed operons mediate cAMP-CRP regulation of pilus-adhesin in *Escherichia coli*. *Molecular Microbiology* 3, 1557–1565.

Goransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelsson, K. and Uhlin, B.E. (1990) Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. *Nature* 344, 682–685.

Hagan, E.C., Lloyd, A.L., Rasko, D.A., Faerber, G.J. and Mobley, H.L. (2010) *Escherichia coli* global gene expression in urine from women with urinary tract infection. *PLoS Pathogens* 6, e1001187.

Hagberg, L., Hull, R., Hull, S., Falkow, S., Freter, R. and Svanborg Eden, C. (1983) Contribution of adhesion to bacterial persistence in the mouse urinary tract. *Infection and Immunity* 40, 265–272.

Hannan, T.J., Totsika, M., Mansfield, K.J., Moore,

K.H., Schembri, M.A. and Hultgren, S.J. (2012) Host–pathogen checkpoints and population bottlenecks in persistent and intracellular uropathogenic *Escherichia coli* bladder infection. *FEMS Microbiology Reviews* 36, 616–648.

Haraoka, M., Hang, L., Frendeus, B., Godaly, G., Burdick, M., Strieter, R. and Svanborg, C. (1999) Neutrophil recruitment and resistance to urinary tract infection. *Journal of Infectious Diseases* 180, 1220–1229.

Hase, K., Kawano, K., Nochi, T., Pontes, G.S., Fukuda, S., Ebisawa, M., Kadokura, K., Tobe, T., Fujimura, Y., Kawano, S., Yabashi, A., Waguri, S., Nakato, G., Kimura, S., Murakami, T., Iimura, M., Hamura, K., Fukuoka, S., Lowe, A.W., Itoh, K., Kiyono, H. and Ohno, H. (2009) Uptake through glycoprotein 2 of FimH(+) bacteria by M cells initiates mucosal immune response. *Nature* 462, 226–230.

Hedlund, M., Wachtler, C., Johansson, E., Hang, L., Somerville, J.E., Darveau, R.P. and Svanborg, C. (1999) P fimbriae-dependent, lipopolysaccharide-independent activation of epithelial cytokine responses. *Molecular Microbiology* 33, 693–703.

Hedlund, M., Frendeus, B., Wachtler, C., Hang, L., Fischer, H. and Svanborg, C. (2001) Type 1 fimbriae deliver an LPS- and TLR4-dependent activation signal to CD14-negative cells. *Molecular Microbiology* 39, 542–552.

Hernday, A.D., Braaten, B.A. and Low, D.A. (2003) The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Molecular Cell* 12, 947–957.

Hernday, A.D., Braaten, B.A., Broitman-Maduro, G., Engelberts, P. and Low, D.A. (2004) Regulation of the pap epigenetic switch by CpxAR: phosphorylated CpxR inhibits transition to the phase ON state by competition with Lrp. *Molecular Cell* 16, 537–547.

Hilbert, D.W., Pascal, K.E., Libby, E.K., Mordechai, E., Adelson, M.E. and Trama, J.P. (2008) Uropathogenic *Escherichia coli* dominantly suppress the innate immune response of bladder epithelial cells by a lipopolysaccharide- and Toll-like receptor 4-independent pathway. *Microbes and Infection* 10, 114–121.

Holden, N.J., Totsika, M., Mahler, E., Roe, A.J., Catherwood, K., Lindner, K., Dobrindt, U. and Gally, D.L. (2006) Demonstration of regulatory cross-talk between P fimbriae and type 1 fimbriae in uropathogenic *Escherichia coli*. *Microbiology* 152, 1143–1153.

Holden, N., Totsika, M., Dixon, L., Catherwood, K. and Gally, D.L. (2007) Regulation of P-fimbrial phase variation frequencies in *Escherichia coli* CFT073. *Infection and Immunity* 75, 3325–3334.

Hooton, T.M. (2012) Clinical practice. Uncomplicated urinary tract infection. *New England Journal of Medicine* 366, 1028–1037.

Hull, R.A., Rudy, D.C., Donovan, W.H., Wieser, I.E., Stewart, C. and Darouiche, R.O. (1999) Virulence properties of *Escherichia coli* 83972, a prototype strain associated with asymptomatic bacteriuria. *Infection and Immunity* 67, 429–432.

Hultgren, S.J., Porter, T.N., Schaeffer, A.J. and Duncan, J.L. (1985) Role of type 1 pili and effects of phase variation on lower urinary tract infections produced by *Escherichia coli*. *Infection and Immunity* 50, 370–377.

Hung, C.S., Bouckaert, J., Hung, D., Pinkner, J., Widberg, C., Defusco, A., Auguste, C.G., Strouse, R., Langermann, S., Waksman, G. and Hultgren, S.J. (2002) Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Molecular Microbiology* 44, 903–915.

Hung, D.L., Raivio, T.L., Jones, C.H., Silhavy, T.J. and Hultgren, S.J. (2001) Cpx signaling pathway monitors biogenesis and affects assembly and expression of P pili. *EMBO Journal* 20, 1508–1518.

Hunstad, D.A., Justice, S.S., Hung, C.S., Lauer, S.R. and Hultgren, S.J. (2005) Suppression of bladder epithelial cytokine responses by uropathogenic *Escherichia coli*. *Infection and Immunity* 73, 3999–4006.

Hvidberg, H., Struve, C., Krogfelt, K.A., Christensen, N., Rasmussen, S.N. and Frimodt-Moller, N. (2000) Development of a long-term ascending urinary tract infection mouse model for antibiotic treatment studies. *Antimicrobial Agents and Chemotherapy* 44, 156–163.

Ikaheimo, R., Siitonen, A., Heiskanen, T., Karkkainen, U., Kuosmanen, P., Lipponen, P. and Makela, P.H. (1996) Recurrence of urinary tract infection in a primary care setting: analysis of a 1-year follow-up of 179 women. *Clinical Infectious Diseases* 22, 91–99.

Jiang, X., Abgottspon, D., Kleeb, S., Rabbani, S., Scharenberg, M., Wittwer, M., Haug, M., Schwardt, O. and Ernst, B. (2012) Antiadhesion therapy for urinary tract infections – a balanced PK/PD profile proved to be key for success. *Journal of Medicinal Chemistry* 55, 4700–4713.

Johanson, I.M., Plos, K., Marklund, B.I. and Svanborg, C. (1993) Pap, papG and prsG DNA sequences in *Escherichia coli* from the fecal flora and the urinary tract. *Microbial Pathogenesis* 15, 121–129.

Johnson, J.R. (1998) papG alleles among

Escherichia coli strains causing urosepsis: associations with other bacterial characteristics and host compromise. *Infection and Immunity* 66, 4568–4571.

Johnson, J.R., Brown, J.J. and Maslow, J.N. (1998a) Clonal distribution of the three alleles of the Gal(alpha1-4)Gal-specific adhesin gene papG among *Escherichia coli* strains from patients with bacteremia. *Journal of Infectious Diseases* 177, 651–661.

Johnson, J.R., Russo, T.A., Brown, J.J. and Stapleton, A. (1998b) papG alleles of *Escherichia coli* strains causing first-episode or recurrent acute cystitis in adult women. *Journal of Infectious Diseases* 177, 97–101.

Johnson, J.R., O'Bryan, T.T., Low, D.A., Ling, G., Delavari, P., Fasching, C., Russo, T.A., Carlino, U. and Stell, A.L. (2000) Evidence of commonality between canine and human extraintestinal pathogenic *Escherichia coli* strains that express papG allele III. *Infection and Immunity* 68, 3327–3336.

Jolley, J.A., Kim, S. and Wing, D.A. (2012) Acute pyelonephritis and associated complications during pregnancy in 2006 in US hospitals. *Journal of Maternal-Fetal and Neonatal Medicine* 25, 2494–2498.

Jones, C.H., Pinkner, J.S., Roth, R., Heuser, J., Nicholes, A.V., Abraham, S.N. and Hultgren, S.J. (1995) FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proceedings of the National Academy of Sciences of the United States of America* 92, 2081–2085.

Jorgensen, I. and Seed, P.C. (2012) How to make it in the urinary tract: a tutorial by *Escherichia coli*. *PLoS Pathogens* 8, e1002907.

Justice, S.S., Hung, C., Theriot, J.A., Fletcher, D.A., Anderson, G.G., Footer, M.J. and Hultgren, S.J. (2004) Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 101, 1333–1338.

Justice, S.S., Hunstad, D.A., Seed, P.C. and Hultgren, S.J. (2006) Filamentation by *Escherichia coli* subverts innate defenses during urinary tract infection. *Proceedings of the National Academy of Sciences of the United States of America* 103, 19884–19889.

Kallenius, G., Mollby, R., Svenson, S.B., Helin, I., Hultberg, H., Cedergren, B. and Winberg, J. (1981) Occurrence of P-fimbriated *Escherichia coli* in urinary tract infections. *Lancet* 2, 1369–1372.

Kaltenbach, L.S., Braaten, B.A. and Low, D.A. (1995) Specific binding of PapI to Lrp-pap DNA complexes. *Journal of Bacteriology* 177, 6449–6455.

Karam, M.R., Oloomi, M., Mahdavi, M., Habibi, M. and Bouzari, S. (2012) Assessment of immune responses of the flagellin (FliC) fused to FimH adhesin of uropathogenic *Escherichia coli*. *Molecular Immunology* 54, 32–39.

Kawamura, T., Vartanian, A.S., Zhou, H. and Dahlquist, F.W. (2011) The design involved in PapI and Lrp regulation of the pap operon. *Journal of Molecular Biology* 409, 311–332.

Keith, B.R., Maurer, L., Spears, P.A. and Orndorff, P.E. (1986) Receptor-binding function of type 1 pili effects bladder colonization by a clinical isolate of *Escherichia coli*. *Infection and Immunity* 53, 693–696.

Kerrn, M.B., Struve, C., Blom, J., Frimodt-Moller, N. and Krogfelt, K.A. (2005) Intracellular persistence of *Escherichia coli* in urinary bladders from mecillinam-treated mice. *Journal of Antimicrobial Chemotherapy* 55, 383–386.

Kisielius, P.V., Schwan, W.R., Amundsen, S.K., Duncan, J.L. and Schaeffer, A.J. (1989) *In vivo* expression and variation of *Escherichia coli* type 1 and P pili in the urine of adults with acute urinary tract infections. *Infection and Immunity* 57, 1656–1662.

Klein, T., Abgottspon, D., Wittwer, M., Rabbani, S., Herold, J., Jiang, X., Kleeb, S., Luthi, C., Scharenberg, M., Bezencon, J., Gubler, E., Pang, L., Smiesko, M., Cutting, B., Schwardt, O. and Ernst, B. (2010) FimH antagonists for the oral treatment of urinary tract infections: from design and synthesis to *in vitro* and *in vivo* evaluation. *Journal of Medicinal Chemistry* 53, 8627–8641.

Klemm, P. (1986) Two regulatory fim genes, fimB and fimE, control the phase variation of type 1 fimbriae in *Escherichia coli*. *EMBO Journal* 5, 1389–1393.

Klemm, P., Roos, V., Ulett, G.C., Svanborg, C. and Schembri, M.A. (2006) Molecular characterization of the *Escherichia coli* asymptomatic bacteriuria strain 83972: the taming of a pathogen. *Infection and Immunity* 74, 781–785.

Klemm, P., Hancock, V. and Schembri, M.A. (2007) Mellowing out: adaptation to commensalism by *Escherichia coli* asymptomatic bacteriuria strain 83972. *Infection and Immunity* 75, 3688–3695.

Klumpp, D.J., Weiser, A.C., Sengupta, S., Forrestal, S.G., Batler, R.A. and Schaeffer, A.J. (2001) Uropathogenic *Escherichia coli* potentiates type 1 pilus-induced apoptosis by suppressing NF-kappaB. *Infection and Immunity* 69, 6689–6695.

Klumpp, D.J., Rycyk, M.T., Chen, M.C., Thumbikat, P., Sengupta, S. and Schaeffer, A.J. (2006)

Uropathogenic *Escherichia coli* induces extrinsic and intrinsic cascades to initiate urothelial apoptosis. *Infection and Immunity* 74, 5106–5113.

Krogfelt, K.A., Bergmans, H. and Klemm, P. (1990) Direct evidence that the FimH protein is the mannose-specific adhesin of *Escherichia coli* type 1 fimbriae. *Infection and Immunity* 58, 1995–1998.

Kuehn, M.J., Heuser, J., Normark, S. and Hultgren, S.J. (1992) P pili in uropathogenic *E. coli* are composite fibres with distinct fibrillar adhesive tips. *Nature* 356, 252–255.

Kuespert, K., Pils, S. and Hauck, C.R. (2006) CEACAMs: their role in physiology and pathophysiology. *Current Opinion in Cell Biology* 18, 565–571.

Kukkonen, M., Raunio, T., Virkola, R., Lahteenmaki, K., Makela, P.H., Klemm, P., Clegg, S. and Korhonen, T.K. (1993) Basement membrane carbohydrate as a target for bacterial adhesion: binding of type I fimbriae of *Salmonella enterica* and *Escherichia coli* to laminin. *Molecular Microbiology* 7, 229–237.

Langermann, S. (1997) Prevention of mucosal *Escherichia coli* infection by FimH-adhesin-based systemic vaccination. *Science* 276, 607–611.

Le Trong, I., Aprikian, P., Kidd, B.A., Forero-Shelton, M., Tchesnokova, V., Rajagopal, P., Rodriguez, V., Interlandi, G., Klevit, R., Vogel, V., Stenkamp, R.E., Sokurenko, E.V. and Thomas, W.E. (2010) Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like beta sheet twisting. *Cell* 141, 645–655.

Leffler, H. and Svanborg-Eden, C. (1981) Glycolipid receptors for uropathogenic *Escherichia coli* on human erythrocytes and uroepithelial cells. *Infection and Immunity* 34, 920–929.

Li, K., Zhou, W., Hong, Y., Sacks, S.H. and Sheerin, N.S. (2009) Synergy between type 1 fimbriae expression and C3 opsonisation increases internalisation of *E. coli* by human tubular epithelial cells. *BMC Microbiology* 9, 64.

Lim, J.K., Gunther, N.W.T., Zhao, H., Johnson, D.E., Keay, S.K. and Mobley, H.L. (1998) *In vivo* phase variation of *Escherichia coli* type 1 fimbrial genes in women with urinary tract infection. *Infection and Immunity* 66, 3303–3310.

Loughman, J.A. and Hunstad, D.A. (2011) Attenuation of human neutrophil migration and function by uropathogenic bacteria. *Microbes and Infection* 13, 555–565.

Loughman, J.A. and Hunstad, D.A. (2012) Induction of indoleamine 2,3-dioxygenase by uropathogenic bacteria attenuates innate responses to epithelial infection. *Journal of Infectious Diseases* 205, 1830–1839.

Malaviya, R., Gao, Z., Thankavel, K., Van Der Merwe, P.A. and Abraham, S.N. (1999) The mast cell tumor necrosis factor alpha response to FimH-expressing *Escherichia coli* is mediated by the glycosylphosphatidylinositol-anchored molecule CD48. *Proceedings of the National Academy of Sciences of the United States of America* 96, 8110–8115.

Martinez, J.J. and Hultgren, S.J. (2002) Requirement of Rho-family GTPases in the invasion of Type 1-piliated uropathogenic *Escherichia coli*. *Cellular Microbiology* 4, 19–28.

Martinez, J.J., Mulvey, M.A., Schilling, J.D., Pinkner, J.S. and Hultgren, S.J. (2000) Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *EMBO Journal* 19, 2803–2812.

McClain, M.S., Blomfield, I.C. and Eisenstein, B.I. (1991) Roles of fimB and fimE in site-specific DNA inversion associated with phase variation of type 1 fimbriae in *Escherichia coli*. *Journal of Bacteriology* 173, 5308–5314.

Melican, K., Sandoval, R.M., Kader, A., Josefsson, L., Tanner, G.A., Molitoris, B.A. and Richter-Dahlfors, A. (2011) Uropathogenic *Escherichia coli* P and Type 1 fimbriae act in synergy in a living host to facilitate renal colonization leading to nephron obstruction. *PLoS Pathogens* 7, e1001298.

Mills, M., Meysick, K.C. and O'Brien, A.D. (2000) Cytotoxic necrotizing factor type 1 of uropathogenic *Escherichia coli* kills cultured human uroepithelial 5637 cells by an apoptotic mechanism. *Infection and Immunity* 68, 5869–5880.

Mo, L., Zhu, X.H., Huang, H.Y., Shapiro, E., Hasty, D.L. and Wu, X.R. (2004) Ablation of the Tamm-Horsfall protein gene increases susceptibility of mice to bladder colonization by type 1-fimbriated *Escherichia coli*. *American Journal of Physiology. Renal Physiology* 286, F795–802.

Mobley, H.L., Jarvis, K.G., Elwood, J.P., Whittle, D.I., Lockatell, C.V., Russell, R.G., Johnson, D.E., Donnenberg, M.S. and Warren, J.W. (1993) Isogenic P-fimbrial deletion mutants of pyelonephritogenic *Escherichia coli*: the role of alpha Gal(1-4) beta Gal binding in virulence of a wild-type strain. *Molecular Microbiology* 10, 143–155.

Mossman, K.L., Mian, M.F., Lauzon, N.M., Gyles, C.L., Lichty, B., Mackenzie, R., Gill, N. and Ashkar, A.A. (2008) Cutting edge: FimH adhesin of type 1 fimbriae is a novel TLR4 ligand. *Journal of Immunology* 181, 6702–6706.

Muller, C.M., Aberg, A., Straseviciene, J., Emody, L., Uhlin, B.E. and Balsalobre, C. (2009) Type 1

fimbriae, a colonization factor of uropathogenic *Escherichia coli*, are controlled by the metabolic sensor CRP-cAMP. *PLoS Pathogens* 5, e1000303.

Mulvey, M.A. (2002) Adhesion and entry of uropathogenic *Escherichia coli*. *Cellular Microbiology* 4, 257–271.

Mulvey, M.A., Lopez-Boado, Y.S., Wilson, C.L., Roth, R., Parks, W.C., Heuser, J. and Hultgren, S.J. (1998) Induction and evasion of host defenses by type 1-piliated uropathogenic *Escherichia coli*. *Science* 282, 1494–1497.

Mulvey, M.A., Schilling, J.D. and Hultgren, S.J. (2001) Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection. *Infection and Immunity* 69, 4572–4579.

Mysorekar, I.U. and Hultgren, S.J. (2006) Mechanisms of uropathogenic *Escherichia coli* persistence and eradication from the urinary tract. *Proceedings of the National Academy of Sciences of the United States of America* 103, 14170–14175.

Nazareth, H., Genagon, S.A. and Russo, T.A. (2007) Extraintestinal pathogenic *Escherichia coli* survives within neutrophils. *Infection and Immunity* 75, 2776–2785.

Newman, E.B. and Lin, R. (1995) Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli*. *Annual Review of Microbiology* 49, 747–775.

Newman, J.V., Burghoff, R.L., Pallesen, L., Krogfelt, K.A., Kristensen, C.S., Laux, D.C. and Cohen, P.S. (1994) Stimulation of *Escherichia coli* F-18Col- type-1 fimbriae synthesis by leuX. *FEMS Microbiology Letters* 122, 281–287.

Norinder, B.S., Koves, B., Yadav, M., Brauner, A. and Svanborg, C. (2012) Do *Escherichia coli* strains causing acute cystitis have a distinct virulence repertoire? *Microbial Pathogenesis* 52, 10–16.

Nou, X., Skinner, B., Braaten, B., Blyn, L., Hirsch, D. and Low, D. (1993) Regulation of pyelonephritis-associated pili phase-variation in *Escherichia coli*: binding of the PapI and the Lrp regulatory proteins is controlled by DNA methylation. *Molecular Microbiology* 7, 545–553.

Ofek, I., Mosek, A. and Sharon, N. (1981) Mannose-specific adherence of *Escherichia coli* freshly excreted in the urine of patients with urinary tract infections, and of isolates subcultured from the infected urine. *Infection and Immunity* 34, 708–711.

O'Gara, J.P. and Dorman, C.J. (2000) Effects of local transcription and H-NS on inversion of the fim switch of *Escherichia coli*. *Molecular Microbiology* 36, 457–466.

O'Hanley, P., Lark, D., Falkow, S. and Schoolnik, G. (1985) Molecular basis of *Escherichia coli* colonization of the upper urinary tract in BALB/c mice. Gal-Gal pili immunization prevents *Escherichia coli* pyelonephritis in the BALB/c mouse model of human pyelonephritis. *Journal of Clinical Investigation* 75, 347–360.

Otto, G., Sandberg, T., Marklund, B.I., Ulleryd, P. and Svanborg, C. (1993) Virulence factors and pap genotype in *Escherichia coli* isolates from women with acute pyelonephritis, with or without bacteremia. *Clinical Infectious Diseases* 17, 448–456.

Palsson-McDermott, E.M. and O'Neill, L.A. (2004) Signal transduction by the lipopolysaccharide receptor, Toll-like receptor-4. *Immunology* 113, 153–162.

Pere, A., Nowicki, B., Saxen, H., Siitonen, A. and Korhonen, T.K. (1987) Expression of P, type-1, and type-1C fimbriae of *Escherichia coli* in the urine of patients with acute urinary tract infection. *Journal of Infectious Diseases* 156, 567–574.

Peterson, S.N. and Reich, N.O. (2006) GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence pap expression. *Journal of Molecular Biology* 355, 459–472.

Peterson, S.N. and Reich, N.O. (2008) Competitive Lrp and Dam assembly at the *pap* regulatory region: implications for mechanisms of epigenetic regulation. *Journal of Molecular Biology* 383, 92–105.

Plos, K., Carter, T., Hull, S., Hull, R. and Svanborg Eden, C. (1990) Frequency and organization of pap homologous DNA in relation to clinical origin of uropathogenic *Escherichia coli*. *Journal of Infectious Diseases* 161, 518–524.

Poggio, T.V., La Torre, J.L. and Scodeller, E.A. (2006) Intranasal immunization with a recombinant truncated FimH adhesin adjuvanted with CpG oligodeoxynucleotides protects mice against uropathogenic *Escherichia coli* challenge. *Canadian Journal of Microbiology* 52, 1093–1102.

Pouttu, R., Puustinen, T., Virkola, R., Hacker, J., Klemm, P. and Korhonen, T.K. (1999) Amino acid residue Ala-62 in the FimH fimbrial adhesin is critical for the adhesiveness of meningitis-associated *Escherichia coli* to collagens. *Molecular Microbiology* 31, 1747–1757.

Pratt, L.A. and Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Molecular Microbiology* 30, 285–293.

Ragnarsdottir, B. and Svanborg, C. (2012) Susceptibility to acute pyelonephritis or

asymptomatic bacteriuria: host–pathogen interaction in urinary tract infections. *Pediatric Nephrology* 27, 2017–2029.

Ragnarsdottir, B., Lutay, N., Gronberg-Hernandez, J., Koves, B. and Svanborg, C. (2011) Genetics of innate immunity and UTI susceptibility. *Nature Reviews. Urology* 8, 449–468.

Roberts, J.A., Hardaway, K., Kaack, B., Fussell, E.N. and Baskin, G. (1984) Prevention of pyelonephritis by immunization with P-fimbriae. *Journal of Urology* 131, 602–607.

Roberts, J.A., Marklund, B.I., Ilver, D., Haslam, D., Kaack, M.B., Baskin, G., Louis, M., Mollby, R., Winberg, J. and Normark, S. (1994) The Gal(alpha 1-4)Gal-specific tip adhesin of *Escherichia coli* P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *Proceedings of the National Academy of Sciences of the United States of America* 91, 11889–11893.

Ronald, A. (2002) The etiology of urinary tract infection: traditional and emerging pathogens. *American Journal of Medicine* 113 Suppl 1A, 14S–19S.

Rosen, D.A., Hooton, T.M., Stamm, W.E., Humphrey, P.A. and Hultgren, S.J. (2007) Detection of intracellular bacterial communities in human urinary tract infection. *PLoS Medicine* 4, e329.

Russo, T.A., Stapleton, A., Wenderoth, S., Hooton, T.M. and Stamm, W.E. (1995) Chromosomal restriction fragment length polymorphism analysis of *Escherichia coli* strains causing recurrent urinary tract infections in young women. *Journal of Infectious Diseases* 172, 440–445.

Samuelsson, P., Hang, L., Wullt, B., Irjala, H. and Svanborg, C. (2004) Toll-like receptor 4 expression and cytokine responses in the human urinary tract mucosa. *Infection and Immunity* 72, 3179–3186.

Sauter, S.L., Rutherfurd, S.M., Wagener, C., Shively, J.E. and Hefta, S.A. (1991) Binding of non-specific cross-reacting antigen, a granulocyte membrane glycoprotein, to *Escherichia coli* expressing type 1 fimbriae. *Infection and Immunity* 59, 2485–2493.

Sauter, S.L., Rutherfurd, S.M., Wagener, C., Shively, J.E. and Hefta, S.A. (1993) Identification of the specific oligosaccharide sites recognized by type 1 fimbriae from *Escherichia coli* on nonspecific cross-reacting antigen, a CD66 cluster granulocyte glycoprotein. *Journal of Biological Chemistry* 268, 15510–15516.

Schaeffer, A.J., Schwan, W.R., Hultgren, S.J. and Duncan, J.L. (1987) Relationship of type 1 pilus expression in *Escherichia coli* to ascending urinary tract infections in mice. *Infection and Immunity* 55, 373–380.

Schappert, S.M. and Rechtsteiner, E.A. (2011) Ambulatory medical care utilization estimates for 2007. *Vital and Health Statistics. Series 13, Data from the National Health Survey*, 1–38.

Schembri, M.A. and Klemm, P. (2001) Biofilm formation in a hydrodynamic environment by novel fimh variants and ramifications for virulence. *Infection and Immunity* 69, 1322–1328.

Schembri, M.A., Sokurenko, E.V. and Klemm, P. (2000) Functional flexibility of the FimH adhesin: insights from a random mutant library. *Infection and Immunity* 68, 2638–2646.

Shin, J.S. (2000) Involvement of cellular caveolae in bacterial entry into mast cells. *Science* 289, 785–788.

Silverblatt, F.J. and Cohen, L.S. (1979) Antipili antibody affords protection against experimental ascending pyelonephritis. *Journal of Clinical Investigation* 64, 333–336.

Silverblatt, F.J., Weinstein, R. and Rene, P. (1982) Protection against experimental pyelonephritis by antibodies to pili. *Scandinavian Journal of Infectious Diseases. Supplementum* 33, 79–82.

Simms, A.N. and Mobley, H.L. (2008) Multiple genes repress motility in uropathogenic *Escherichia coli* constitutively expressing type 1 fimbriae. *Journal of Bacteriology* 190, 3747–3756.

Snyder, J.A., Haugen, B.J., Buckles, E.L., Lockatell, C.V., Johnson, D.E., Donnenberg, M.S., Welch, R.A. and Mobley, H.L. (2004) Transcriptome of uropathogenic *Escherichia coli* during urinary tract infection. *Infection and Immunity* 72, 6373–6381.

Snyder, J.A., Haugen, B.J., Lockatell, C.V., Maroncle, N., Hagan, E.C., Johnson, D.E., Welch, R.A. and Mobley, H.L. (2005) Coordinate expression of fimbriae in uropathogenic *Escherichia coli*. *Infection and Immunity* 73, 7588–7596.

Sokurenko, E.V., Courtney, H.S., Abraham, S.N., Klemm, P. and Hasty, D.L. (1992) Functional heterogeneity of type 1 fimbriae of *Escherichia coli*. *Infection and Immunity* 60, 4709–4719.

Sokurenko, E.V., Courtney, H.S., Maslow, J., Siitonen, A. and Hasty, D.L. (1995) Quantitative differences in adhesiveness of type 1 fimbriated *Escherichia coli* due to structural differences in fimH genes. *Journal of Bacteriology* 177, 3680–3686.

Sokurenko, E.V., Chesnokova, V., Doyle, R.J. and Hasty, D.L. (1997) Diversity of the *Escherichia*

coli type 1 fimbrial lectin. Differential binding to mannosides and uroepithelial cells. *Journal of Biological Chemistry* 272, 17880–17886.

Sokurenko, E.V., Chesnokova, V., Dykhuizen, D.E., Ofek, I., Wu, X.R., Krogfelt, K.A., Struve, C., Schembri, M.A. and Hasty, D.L. (1998) Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proceedings of the National Academy of Sciences of the United States of America* 95, 8922–8926.

Southgate, J., Kennedy, W., Hutton, K.A. and Trejdosiewicz, L.K. (1995) Expression and *in vitro* regulation of integrins by normal human urothelial cells. *Cell Adhesion and Communication* 3, 231–242.

Stickler, D.J. (2008) Bacterial biofilms in patients with indwelling urinary catheters. *Nature Clinical Practice. Urology* 5, 598–608.

Stromberg, N., Marklund, B.I., Lund, B., Ilver, D., Hamers, A., Gaastra, W., Karlsson, K.A. and Normark, S. (1990) Host-specificity of uropathogenic *Escherichia coli* depends on differences in binding specificity to Gal alpha 1-4Gal-containing isoreceptors. *EMBO Journal* 9, 2001–2010.

Taganna, J., De Boer, A.R., Wuhrer, M. and Bouckaert, J. (2011) Glycosylation changes as important factors for the susceptibility to urinary tract infection. *Biochemical Society Transactions* 39, 349–354.

Tewari, R., MacGregor, J.I., Ikeda, T., Little, J.R., Hultgren, S.J. and Abraham, S.N. (1993) Neutrophil activation by nascent FimH subunits of type 1 fimbriae purified from the periplasm of *Escherichia coli*. *Journal of Biological Chemistry* 268, 3009–3015.

Thanassi, D.G., Saulino, E.T., Lombardo, M.J., Roth, R., Heuser, J. and Hultgren, S.J. (1998) The PapC usher forms an oligomeric channel: implications for pilus biogenesis across the outer membrane. *Proceedings of the National Academy of Sciences of the United States of America* 95, 3146–3151.

Thanassi, D.G., Bliska, J.B. and Christie, P.J. (2012) Surface organelles assembled by secretion systems of Gram-negative bacteria: diversity in structure and function. *FEMS Microbiology Reviews* 36, 1046–1082.

Thomas, W.E., Trintchina, E., Forero, M., Vogel, V. and Sokurenko, E.V. (2002) Bacterial adhesion to target cells enhanced by shear force. *Cell* 109, 913–923.

Thomas, W.E., Nilsson, L.M., Forero, M., Sokurenko, E.V. and Vogel, V. (2004) Shear-dependent 'stick-and-roll' adhesion of type 1 fimbriated *Escherichia coli*. *Molecular Microbiology* 53, 1545–1557.

Thumbikat, P., Berry, R.E., Zhou, G., Billips, B.K., Yaggie, R.E., Zaichuk, T., Sun, T.T., Schaeffer, A.J. and Klumpp, D.J. (2009) Bacteria-induced uroplakin signaling mediates bladder response to infection. *PLoS Pathogens* 5, e1000415.

Visvikis, O., Boyer, L., Torrino, S., Doye, A., Lemonnier, M., Lores, P., Rolando, M., Flatau, G., Mettouchi, A., Bouvard, D., Veiga, E., Gacon, G., Cossart, P. and Lemichez, E. (2011) *Escherichia coli* producing CNF1 toxin hijacks Tollip to trigger Rac1-dependent cell invasion. *Traffic* 12, 579–590.

Wang, H., Min, G., Glockshuber, R., Sun, T.T. and Kong, X.P. (2009) Uropathogenic *E. coli* adhesin-induced host cell receptor conformational changes: implications in transmembrane signaling transduction. *Journal of Molecular Biology* 392, 352–361.

Wang, Z., Humphrey, C., Frilot, N., Wang, G., Nie, Z., Moniri, N.H. and Daaka, Y. (2011) Dynamin2- and endothelial nitric oxide synthase-regulated invasion of bladder epithelial cells by uropathogenic *Escherichia coli*. *Journal of Cell Biology* 192, 101–110.

Weissman, S.J., Chattopadhyay, S., Aprikian, P., Obata-Yasuoka, M., Yarova-Yarovaya, Y., Stapleton, A., Ba-Thein, W., Dykhuizen, D., Johnson, J.R. and Sokurenko, E.V. (2006) Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. *Molecular Microbiology* 59, 975–988.

Weyand, N.J., Braaten, B.A., Van Der Woude, M., Tucker, J. and Low, D.A. (2001) The essential role of the promoter-proximal subunit of CAP in pap phase variation: Lrp- and helical phase-dependent activation of *papBA* transcription by CAP from -215. *Molecular Microbiology* 39, 1504–1522.

White-Ziegler, C.A., Angus Hill, M.L., Braaten, B.A., Van Der Woude, M.W. and Low, D.A. (1998) Thermoregulation of *Escherichia coli pap* transcription: H-NS is a temperature-dependent DNA methylation blocking factor. *Molecular Microbiology* 28, 1121–1137.

Wiles, T.J., Dhakal, B.K., Eto, D.S. and Mulvey, M.A. (2008) Inactivation of host Akt/protein kinase B signaling by bacterial pore-forming toxins. *Molecular Biology of the Cell* 19, 1427–1438.

Wold, A.E., Thorssen, M., Hull, S. and Eden, C.S. (1988) Attachment of *Escherichia coli* via mannose- or Gal alpha 1–4Gal beta-containing receptors to human colonic epithelial cells. *Infection and Immunity* 56, 2531–2537.

Wold, A.E., Mestecky, J., Tomana, M., Kobata, A., Ohbayashi, H., Endo, T. and Eden, C.S. (1990) Secretory immunoglobulin A carries oligosaccharide receptors for *Escherichia coli* type 1 fimbrial lectin. *Infection and Immunity* 58, 3073–3077.

Wright, K.J., Seed, P.C. and Hultgren, S.J. (2007) Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cellular Microbiology* 9, 2230–2241.

Wu, X.R., Kong, X.P., Pellicer, A., Kreibich, G. and Sun, T.T. (2009) Uroplakins in urothelial biology, function, and disease. *Kidney International* 75, 1153–1165.

Wullt, B. (2003) The role of P fimbriae for *Escherichia coli* establishment and mucosal inflammation in the human urinary tract. *International Journal of Antimicrobial Agents* 21, 605–621.

Wullt, B., Bergsten, G., Connell, H., Rollano, P., Gebretsadik, N., Hull, R. and Svanborg, C. (2000) P fimbriae enhance the early establishment of *Escherichia coli* in the human urinary tract. *Molecular Microbiology* 38, 456–464.

Wullt, B., Bergsten, G., Connell, H., Rollano, P., Gebratsedik, N., Hang, L. and Svanborg, C. (2001) P-fimbriae trigger mucosal responses to *Escherichia coli* in the human urinary tract. *Cellular Microbiology* 3, 255–264.

Xia, Y., Gally, D., Forsman-Semb, K. and Uhlin, B.E. (2000) Regulatory cross-talk between adhesin operons in *Escherichia coli*: inhibition of type 1 fimbriae expression by the PapB protein. *EMBO Journal* 19, 1450–1457.

Yu, S. and Lowe, A.W. (2009) The pancreatic zymogen granule membrane protein, GP2, binds *Escherichia coli* Type 1 fimbriae. *BMC Gastroenterology* 9, 58.

Zhou, G., Mo, W.J., Sebbel, P., Min, G., Neubert, T.A., Glockshuber, R., Wu, X.R., Sun, T.T. and Kong, X.P. (2001) Uroplakin Ia is the urothelial receptor for uropathogenic *Escherichia coli*: evidence from *in vitro* FimH binding. *Journal of Cell Science* 114, 4095–4103.

CHAPTER 3

THE EXTRAINTESTINAL PATHOGENIC ESCHERICHIA COLI FACTOR CONSTRAINS THE GENOTOXIC EFFECTS OF THE RECQ-LIKE HELICASE RQLH

Reprinted with permission from Public Library of Science

Russell, CW, Mulvey MA. (2015) The Extraintestinal Pathogenic *Escherichia coli* Factor RqlI Constrains the Genotoxic Effects of the RecQ-Like Helicase RqlH. PLoS Pathog. 11(12):e1005317. doi: 10.1371/journal.ppat.1005317.

PLOS | PATHOGENS

RESEARCH ARTICLE

The Extraintestinal Pathogenic *Escherichia coli* Factor RqlI Constrains the Genotoxic Effects of the RecQ-Like Helicase RqlH

Colin W. Russell, Matthew A. Mulvey*

University of Utah School of Medicine, Department of Pathology, Division of Microbiology and Immunology, Salt Lake City, Utah, United States of America

* mulvey@path.utah.edu

Citation: Russell CW, Mulvey MA (2015) The Extraintestinal Pathogenic *Escherichia coli* Factor RqlI Constrains the Genotoxic Effects of the RecQ-Like Helicase RqlH. PLoS Pathog 11(12): e1005317. doi:10.1371/journal.ppat.1005317

Editor: Tomoko Kubori, Osaka University, JAPAN

Received: August 5, 2015

Accepted: November 10, 2015

Published: December 4, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Research in the Mulvey lab was funded by grant AI095647 from the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Extraintestinal pathogenic *Escherichia coli* colonize the human gut and can spread to other body sites to induce diseases such as urinary tract infections, sepsis, and meningitis. A complete understanding of the infection process is hindered by both the inherent genetic diversity of *E. coli* and the large number of unstudied genes. Here, we focus on the uncharacterized gene *rqlI*, which our lab recently uncovered in a Tn-seq screen for bacterial genes required within a zebrafish model of infection. We demonstrate that the Δ*rqlI* mutant experiences a growth defect and increased DNA stress in low oxygen conditions. In a genetic screen for suppressor mutations in the Δ*rqlI* strain, we found that the shortcomings of the Δ*rqlI* mutant are attributable to the activity of RqlH, which is known in other bacteria to be a helicase of the RecQ family that contains a phosphoribosyltransferase (PRTase) domain. Disruption of *rqlH* rescues the Δ*rqlI* strain in both *in vivo* and *in vitro* assays, while the expression of RqlH alone activates the SOS response coincident with bacterial filamentation, heightened sensitivity to DNA damage, and an increased mutation rate. The analysis of truncation mutants indicates that, in the absence of RqlI, RqlH toxicity is due to its PRTase domain. Complementary studies demonstrate that the toxicity of RqlH is modulated in a context-dependent fashion by overlapping domains within RqlI. This regulation is seemingly direct, given that the two proteins physically interact and form an operon. Interestingly, RqlH and RqlI orthologs are encoded by a diverse group of bacteria, but in many of these microbes, and especially in Gram-positive organisms, *rqlH* is found in the absence of *rqlI*. In total, this work shows that RqlH and RqlI can act in a strain-specific fashion akin to a toxin-antitoxin system in which toxicity is mediated by an atypical helicase-associated PRTase domain.

Author Summary

Extraintestinal pathogenic *Escherichia coli* (ExPEC) cause the majority of urinary tract infections, and are also able to infect the bloodstream, meninges, and various other sites within the human host. These infections are becoming increasingly difficult to treat as

ExPEC strains gain resistance to many of the antibiotics that are commonly used in the clinic. The development of improved treatment strategies requires a deeper understanding of the factors that promote ExPEC fitness and virulence within the host. In genetic screens, we identified a functionally uncharacterized protein, RqlI, which promotes ExPEC survival within diverse host environments. We find that RqlI binds to and works in tandem with RqlH, a protein that has been shown in other bacteria to unwind DNA. In the absence of RqlI, we found that RqlH can become toxic to ExPEC, causing DNA damage and slower growth. A specific part of RqlH that is predicted to manipulate the nucleotides that make up DNA is responsible for this toxicity. The ability of RqlH to inhibit bacterial growth when not held in check by RqlI suggests that the specific inactivation of RqlI could have therapeutic value in combating ExPEC and other pathogens that express these proteins.

Introduction

Extraintestinal pathogenic *Escherichia coli* (ExPEC) comprise a group of *E. coli* strains that harmlessly reside within the gastrointestinal tract, but that are capable of infecting extraintestinal niches such as the urinary tract, bloodstream, and meninges [1,2]. Urinary tract infections (UTIs) alone constitute a major global economic burden due to the high incidence rate of UTIs. Indeed, more than half of all women will experience a UTI during their lifetime [3], with bladder infections being responsible for more than $2 billion in medical care costs in the United States annually [4]. Additionally, *E. coli* is the leading cause of bacteremia [5], and is especially problematic in high risk, immunocompromised populations such as the elderly [6,7] and cancer patients undergoing chemotherapy [8]. *E. coli* is also the leading cause of death in early-onset neonatal sepsis [9]. The burden imposed by ExPEC infection is expanding as the prevalence of antibiotic resistant strains increases [10], highlighting the need for a deeper understanding of ExPEC-mediated pathogenesis.

A major hurdle in deciphering the ExPEC life cycle is the genetic heterogeneity inherent in these strains. Each ExPEC isolate contains ~5,000 genes, of which about half comprise a core genome that is shared with most other *E. coli* strains [11]. The genes outside of the core genome are much less conserved and account for the bulk of the genomic heterogeneity observed among ExPEC isolates. This flexible genome is in large part a consequence of horizontal gene transfer among ExPEC strains and other bacteria. The ExPEC pan-genome, which includes all distinct genes found within ExPEC isolates, contains at least 14,877 genes [11]. Due to the dynamic nature of the pan-genome, this estimate is likely to increase as more isolates are sequenced [12]. Because of the genetic differences from one ExPEC isolate to another it is difficult to define a specific set of genes that are universally required for ExPEC fitness.

Complicating matters further is the fact that many ExPEC-associated genes are functionally undefined. One method for dealing with the large number of uncharacterized genes is to employ high-throughput analyses that can assay the role of many genes in a given condition simultaneously. One especially useful technique is Tn-seq, in which a pool of transposon mutants is placed under a selective pressure, and the quantity of individual mutant variants before and after the selection are enumerated by next-generation sequencing [13]. Despite their utility, these types of analyses often result in lists of genes that are required for fitness under a given condition without yielding clear insight into the actual functions of individual genes. For example, in a recent Tn-seq screen for factors that affect ExPEC fitness in a zebrafish infection model, many of the 981 genetic elements that were found to be important for host colonization had never been studied [14]. Annotations for 152 of the 981 loci (15.5%) that

were identified as required for zebrafish colonization include the words "hypothetical" or "putative". In addition, given the limitations of the current pipelines used to annotate genes, many more of the identified loci are likely misannotated [15]. Thus, understanding why a gene is important for fitness under a given condition can be complicated. This problem can be partially remedied by comparing overlapping datasets from high-throughput experiments carried out in various defined conditions [16,17]. This strategy can help uncover roles for previously undefined genes, but additional detailed analysis is still required to validate the high-throughput data and to unambiguously specify gene function.

In our recent Tn-seq screen using the ExPEC cystitis isolate F11, we identified the hypothetical gene *EcF11_3933* as an important mediator of pathogen fitness during both localized and systemic infections within the surrogate zebrafish host [14]. This gene, which we propose to rename RecQ-Like Helicase Interactor (*rqlI*) for reasons that will become clear, was also critical to the ability of F11 to colonize the mouse urinary tract and for pathogen survival in a mouse model of sepsis. Considering the importance of *rqlI* to pathogen survival in diverse hosts and host environments, we set out to define the specific function of RqlI. Here we report that RqlI is especially critical for pathogen growth under low oxygen conditions, as found within many host niches. In the absence of RqlI, the bacteria have higher mutation rates, increased activation of the SOS stress response system, and decreased growth. These effects are remedied by deletion of *rqlH*, a gene immediately upstream of *rqlI* that encodes RecQ-like Helicase (RqlH). Deletion of *rqlH* also rescues the *in vivo* survival defects observed with the Δ*rqlI* mutant in several mouse models. Genetic and biochemical analyses indicate that a function of RqlI within ExPEC is to limit the genotoxic effects of RqlH, which we mapped to an unusual phosphoribosyltransferase (PRTase) domain localized at the C-terminus of RqlH. Interestingly, we find that RqlH and RqlI homologs are encoded by a diverse collection of bacterial species, though the *rqlH* and *rqlI* genes are not always paired as in F11. These observations suggest that RqlI can have both RqlH-dependent and RqlH-independent functions, and that these two proteins have important roles in the life cycles of a wide range of bacteria.

Results

The Δ*rqlI* mutant exhibits a growth defect in microaerobic conditions

An initial analysis of the RqlI protein revealed the presence of three putative domains (Fig 1A). First, RqlI is predicted to contain a molybdenum cofactor (MoCo) binding domain, which is often found in proteins that facilitate metabolism in low oxygen environments [18]. For example, MoCo facilitates redox reactions critical for using alternative terminal electron acceptors when oxygen is scarce. Second, overlapping the MoCo binding domain is a region with homology to the DNA-protecting protein DprA, which has been shown to be important for the uptake of exogenous DNA in naturally competent bacteria [19]. Third, at the C-terminus of RqlI is a helix-turn-helix motif predicted to bind DNA.

Given the involvement of MoCo in anaerobic respiration [18], and given the putative MoCo-binding domain contained within RqlI, we disrupted the *rqlI* gene in the reference ExPEC strain F11 and tested the ability of the mutant to multiply in varying oxygen levels. When grown for 24 hours in a modified M9 minimal media exposed to atmospheric oxygen (~20% O_2), F11Δ*rqlI* exhibits a slight growth defect in comparison to the wild type (WT) strain (Fig 1B). However, in microaerobic culture conditions (6–12% O_2) growth of the Δ*rqlI* mutant was markedly worse than WT (Fig 1B). This defect was rescued by complementation with the RqlI expression plasmid pCWR16. Furthermore, we found that the exponential phase generation time of the Δ*rqlI* mutant was very similar to WT cells under aerobic conditions, but was 3-fold greater than WT bacteria in a microaerobic environment (Fig 1C). Under microaerobic

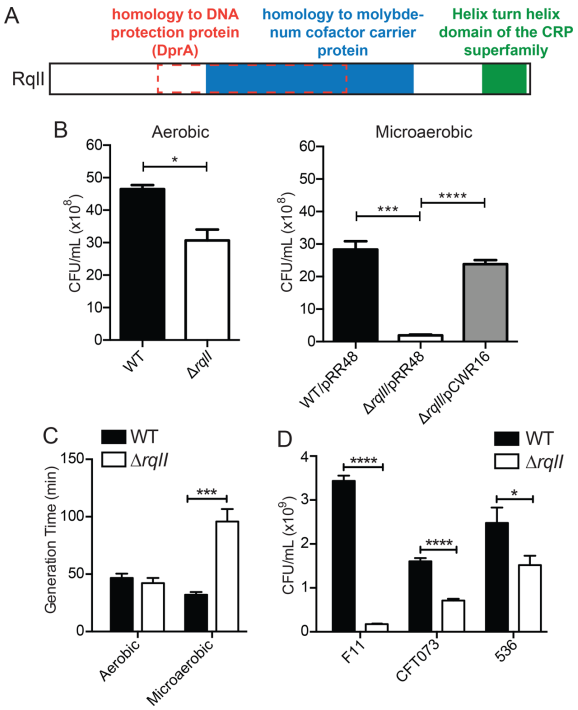


Fig 1. Growth of *rqlI* deletion mutants is severely impaired in low oxygen conditions. (A) A schematic of predicted domains and homology regions of RqlI. (B) WT F11 or F11Δ*rqlI* were grown either aerobically or microaerobically for 24 h in a modified M9 media, at which point the cultures were titered. Under microaerobic conditions, strains that carried either an IPTG-inducible *rqlI* expression vector (pCWR16) or the empty vector control (pRR48) were grown in the presence of IPTG. (C) WT F11 or F11Δ*rqlI* were grown either aerobically or microaerobically, and the cultures were titered at two different time points during exponential phase in order to calculate generation times. (D) The indicated WT and Δ*rqlI* mutant ExPEC isolates were grown microaerobically for 24 h and titered. Bars indicate mean values ± SEM from three or more independent experiments performed in triplicate. *, $P \leq 0.05$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$; as determined by Student's *t* test.

doi:10.1371/journal.ppat.1005317.g001

conditions, deletion of *rqlI* also limited the growth of two other reference ExPEC strains, CFT073 and 536 (Fig 1D).

The SOS response is activated in the Δ*rqlI* mutant during microaerobic growth

We examined the general morphology of F11Δ*rqlI* using light microscopy and noticed that the mutant cells were often longer (more filamentous) than WT bacteria (Fig 2A–2D). The elongated Δ*rqlI* mutant cells were especially abundant, and reached greater lengths, under microaerobic or anaerobic conditions (Fig 2B–2D). One reason for bacterial cell length to increase is the induction of the SOS response, which is initiated when DNA damage is sensed [20]. The

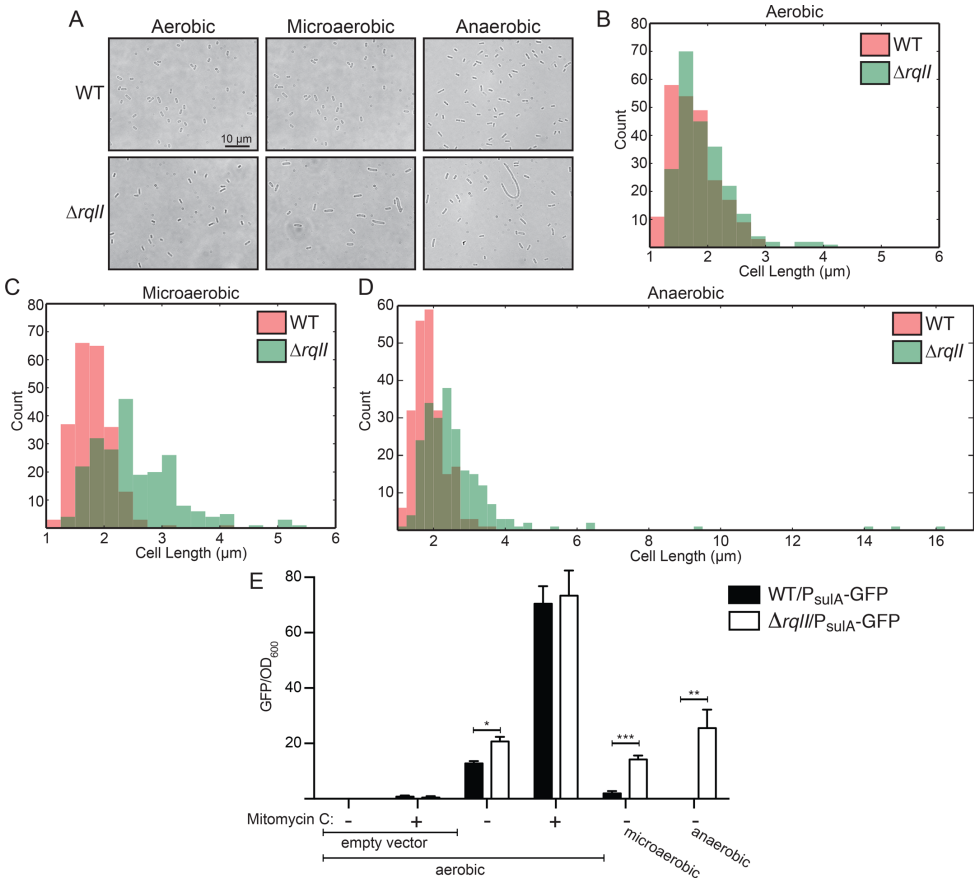


Fig 2. The *rqlI* mutant experiences increased genotoxic stress, especially under low oxygen conditions. (A-D) WT F11 or F11Δ*rqlI* were grown 24 h aerobically, microaerobically, or anaerobically, at which point bacteria were spread on a slide and imaged. (A) Representative images. (B-D) Histograms show the distribution of bacterial cell lengths after growth at varying oxygen levels. (E) WT F11 and F11Δ*rqlI* strains carrying either the P_{sulA}-GFP reporter plasmid (pJLJ3) or an empty vector control (pJLJ1) were grown aerobically ± mitomycin C, microaerobically, or anaerobically, as indicated. Fluorescence measurements were normalized to culture densities. Bars indicate mean values ± SEM from three or more independent experiments performed in triplicate. *, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; as determined by Student's *t* test.

doi:10.1371/journal.ppat.1005317.g002

SOS response involves the upregulation of *sulA* expression, which functions to temporarily inhibit cell division while DNA repair occurs. Using a P_{sulA}-GFP fluorescent reporter construct, we found that F11Δ*rqlI* expressed markedly higher levels of *sulA* than the WT strain in aerobic, microaerobic, and anaerobic conditions (Fig 2E). The differences in *sulA* expression levels between WT and the Δ*rqlI* mutant grew more pronounced with decreasing amounts of oxygen in the culture. In control samples, *sulA* expression levels in the WT and Δ*rqlI* strains were

similarly increased upon treatment with the DNA-damaging agent mitomycin C. These fluorescent measurements were confirmed by western blots probed with GFP-specific antibody (S1 Fig). In total, these data demonstrate that in the absence of RqlI there is increased activation of the SOS response, coincident with an oxygen-sensitive decrease in bacterial replication.

RqlH provokes the SOS response and inhibits bacterial growth in the absence of RqlI

In order to understand why *sulA* expression is increased in the Δ*rqlI* strain, a screen was designed to search for factors that promote SOS induction in the absence of RqlI (Fig 3A). An F11Δ*rqlI*Δ*lacZY* strain was transformed with plasmid pCWR2 containing the *lacZ* gene driven by the *sulA* promoter, providing a convenient readout of *sulA* expression levels on tetrazolium lactose plates. On these plates, high LacZ expression—reflecting high *sulA* promoter activity—results in colonies with pink to white color, while low LacZ activity due to low *sulA* promoter activity results in a red colony color. F11Δ*rqlI*Δ*lacZY*/pCWR2 was randomly mutagenized using the *mariner* transposon from pSAM_Ec [14], and then plated onto selective tetrazolium lactose plates. Approximately 20,000 colonies were screened for darker colony color, indicative of a decreased SOS response. Transposon insertion locations were mapped in potential suppressor mutants, and six independent insertions were identified within *rqlH* (Fig 3B). This gene encodes a homologue of the *Mycobacterium smegmatis* RqlH protein (MSMEG_5935), a RecQ-Like Helicase that is known to unwind dsDNA in a 3' to 5' orientation [21]. The F11 and *M*. *smegmatis* RqlH proteins align along their entire lengths, with 41.2% identity (55.7% similarity) overall (S2 Fig). Interestingly—and perhaps, in retrospect, not surprisingly—the F11 *rqlH* gene is also found adjacent to *rqlI* on the genome (Fig 4A). To confirm the results of the screen, *rqlH* and *rqlI* were deleted from F11, separately and in combination, and P_{sulA}-*GFP* expression levels were measured under low oxygen conditions (Fig 3C). As before, GFP expression levels in the Δ*rqlI* strain were higher than in WT F11, whereas strains lacking only *rqlH* or both *rqlH* and *rqlI* had GFP levels equal to that of WT. These data indicate that in the absence of RqlI, RqlH can trigger the SOS response, resulting in increased *sulA* expression.

To test whether the reduction in growth of F11Δ*rqlI* under microaerobic conditions could also be attributed to RqlH activity, WT F11 and the Δ*rqlI*, Δ*rqlH*, and Δ*rqlHI* strains were grown in low oxygen cultures. Relative to the WT strain, only the Δ*rqlI* mutant grew poorly in these assays (Fig 3D). In addition, we found that overexpression of RqlH caused a marked reduction in bacterial growth under both aerobic and microaerobic conditions (S3 Fig). This effect was seen regardless of genetic background, but was more pronounced in strains that lacked RqlI. Together, these results indicate that RqlH activity in the absence of RqlI both induces the SOS response and inhibits bacterial growth.

RqlH increases bacterial sensitivity to DNA damage in the absence of RqlI

As the SOS response is activated by DNA damage and is responsible for initiating DNA repair programs, we examined the abilities of the *rqlI* and *rqlH* mutants to deal with exogenous DNA damage. WT F11 and mutant derivatives were treated with the DNA crosslinking agent mitomycin C and surviving bacterial cells were quantified. The Δ*rqlI* mutant had increased sensitivity to mitomycin C, a defect that was again rescued in the Δ*rqlHI* double mutant strain (Fig 3E) and by complementation with the RqlI expression plasmid pCWR16 (Fig 3F). A similar trend was observed when the cells were treated with UV light. The Δ*rqlI* mutant was more sensitive to both 25 and 50 J/m^2 of UV treatment, although this effect was only statistically significant at 25 J/m^2 (Fig 3G). In contrast, the Δ*rqlH* and Δ*rqlHI* mutants survived UV exposure at levels

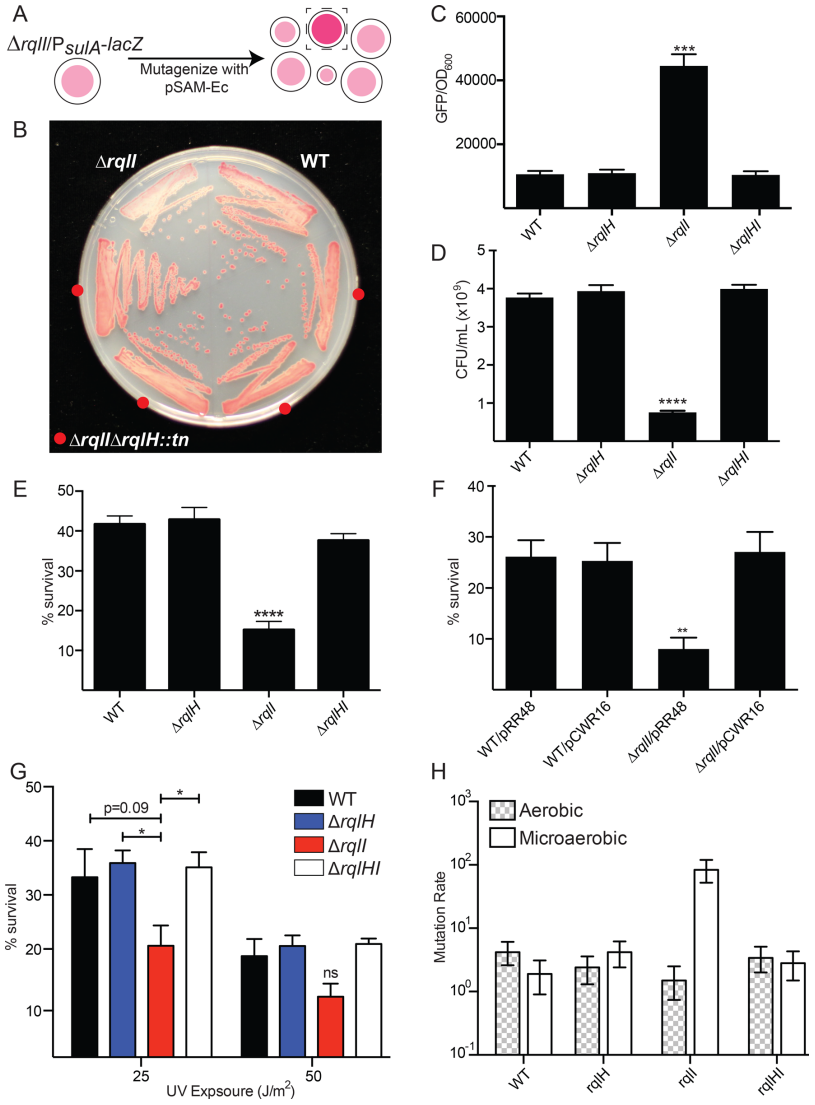


Fig 3. Mutation of *rqlH* rescues the *rqlI* mutant defect *in vitro*. (A) A schematic of the screen used to search for suppressor mutations of the Δ*rqlI* mutant. Bacteria carrying a P_{sulA}-*lacZ* reporter plasmid (pCWR2) were transposon-mutagenized and screened on tetrazolium lactose agar for suppressor mutants with decreased *sulA* promoter activity, (distinguishable by darker colony color). (B) A tetrazolium lactose plate demonstrating the colony colors for WT, F11Δ*rqlI*, and four independent F11Δ*rqlI*Δ*rqlH*::*tn* transposon mutants that were isolated from the screen. (C) Bacteria carrying pJLJ3 (P_{sulA}-GFP) were grown microaerobically and GFP levels and culture densities were measured to quantify *sulA* expression. (D) Microaerobic growth of WT, Δ*rqlH*, Δ*rqlI*, and

Δ*rqlHI* strains, as measured by dilution plating of 24 h cultures. (E) Bacteria were grown aerobically for 4 h, at which point mitomycin C was added to a final concentration of 0.25 μg/mL for 1 h. The cultures were titered immediately before and 1 h after addition of mitomycin C, and survival was calculated. (F) WT and Δ*rqlI* cells carrying either an empty vector control (pRR48), or the RqlI expression plasmid pCWR16 were grown and tested for mitomycin C survival as in (E). (G) Bacteria were grown aerobically for 4 h, then diluted and spread onto plates to get ~100 colonies per plate. The plates were then exposed to various levels of UV radiation and incubated overnight, at which time percent survival was calculated. In C-G, bars indicate mean values ± SEM from three or more independent experiments performed in triplicate. *, $P \leq 0.05$; ***, $P \leq 0.001$; ****, $P \leq 0.0001$ in comparison with WT; as determined by Student's *t* tests. (H) Bacteria were grown either for 4 h aerobically or 24 h microaerobically. Aliquots from each culture were used to titer viable bacteria, and the remaining cultures were plated onto agar containing nalidixic acid. The numbers of nalidixic acid resistant mutants recovered were used as input for FALCOR to calculate mutation rates (number of mutations per 10^{10} cells per generation). Bars indicate mean values ± 95% confidence intervals.

doi:10.1371/journal.ppat.1005317.g003

similar to the WT cells. These data indicate that the Δ*rqlI* mutant is more sensitive to DNA damage in a manner that is dependent on RqlH.

To determine if the increased sensitivity of the Δ*rqlI* mutant to exogenous DNA damaging agents correlates with a decreased ability to deal with endogenous DNA damage, spontaneous mutation rates that resulted in nalidixic acid resistance were quantified using fluctuation assays [22]. WT F11 and the Δ*rqlI*, Δ*rqlH* and Δ*rqlHI* strains all exhibited low mutation rates of about 3 mutations per 10^{10} cells per generation when grown under aerobic conditions (Fig 3H). Growth under microaerobic conditions did not change the mutation rates for WT F11, F11Δ*rqlH*, or F11Δ*rqlHI*, but resulted in nearly an 80-fold increase in the mutation rate of the Δ*rqlI* mutant. These data suggest that in the absence of RqlI, RqlH makes ExPEC more sensitive to DNA damage and the accumulation of spontaneous mutations.

rqlH and *rqlI* are co-expressed in an oxygen-insensitive fashion

Since the *rqlH* and *rqlI* genes are located immediately adjacent to each other on the genome, but are separated from upstream and downstream genes, it is likely that they form a two-gene operon (Fig 4A). To test this idea, we performed RT-PCR using primer sets spanning the *rqlH* and *rqlI* loci, as well as one set internal to *rqlI*, as depicted in Fig 4A. Products expected if *rqlH* and *rqlI* are co-transcribed as part of the same transcript were obtained from F11 cDNA, but were not seen in control reactions that lacked reverse transcriptase (Fig 4B). These results indicate that *rqlH* and *rqlI* form an operon.

Given that the Δ*rqlI* mutant defect is exacerbated by microaerobic conditions, we tested whether *rqlHI* transcription is modulated by oxygen levels. Using primers that are specific for *rqlH*, for *rqlI*, or that span *rqlHI*, we were unable to detect any changes in *rqlHI* expression in microaerobic growth as compared to aerobic growth (Fig 4C). In contrast, we observed that transcription of *frdA*, which is sensitive to oxygen levels [23], was significantly upregulated in F11 grown under microaerobic conditions.

The PRTase domain of RqlH is genotoxic in the absence of RqlI

RqlH contains an N-terminal helicase domain as well as a C-terminal phosphoribosyltransferase (PRTase) domain that is unique among known helicases [21]. The functional importance of the PRTase domain is unclear, though it is dispensable for the ATPase and helicase activities of RqlH in *M. smegmatis* [21]. To define domain(s) within RqlH that are responsible for the genotoxic effects observed in the absence of RqlI, we expressed epitope-tagged RqlH and several mutant variants in the K-12 strain MG1655, which lacks both RqlH and RqlI. Expression of each RqlH variant was confirmed by western blot (S4 Fig). The recombinant MG1655 strains were diluted into modified M9 minimal media and grown for 6 h with aeration, at which point the optical density (OD_{600}) of each culture and the numbers of viable bacteria (CFU/mL) were measured. Expression of the WT RqlH significantly reduced both the growth of MG1655 and the numbers of CFUs recovered (Figs 5 and S5). RqlH expression also stimulated the formation

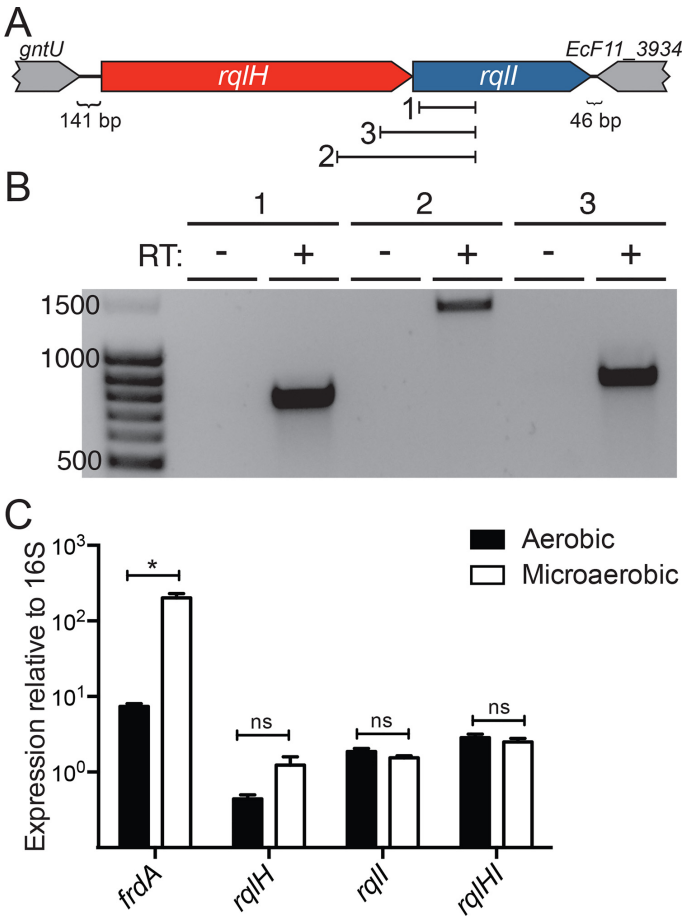
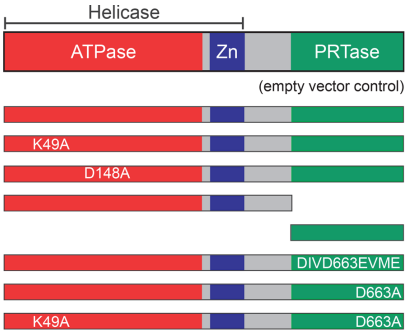


Fig 4. *rqlH* and *rqlI* are co-expressed and are not regulated by oxygen levels. (A) A schematic of the *rqlHI* operon and surrounding genes, with the three primer sets used in (B) indicated. (B) RNA was isolated from F11 WT bacteria growing aerobically, and cDNA was made and used as template in PCRs using the three primer sets indicated in (A). Each primer set was used with a sample lacking reverse transcriptase (RT) as a control. Ladder band sizes are indicated at left. (C) F11 WT bacteria were grown either aerobically or microaerobically, RNA was isolated, and cDNA was made. Several primer sets were used in qRT-PCR reactions to measure levels of *frdA*, *rqlH*, *rqlI*, and *rqlHI* transcripts under the two oxygen conditions. All expression values were normalized to 16S rRNA. Bars indicate mean values ± SEM from three or more independent experiments performed in triplicate. *, $P \leq 0.05$; as determined by Student's *t* test.

doi:10.1371/journal.ppat.1005317.g004



Protein	Plasmid	OD_{600}	CFU/mL	Cell Lengths	*sulA*-GFP
None	pBAD33	1.00	1.00	WT	No
RqlH	pCWR21	0.09 #	0.01 #	Long #	Yes #
K49A	pCWR23	0.98	0.05	Long #	Yes #
D148A	pCWR24	1.00	0.02	Long #	Yes #
1-509	pCWR25	1.09	1.00	WT	No
509-698	pCWR26	0.96	0.98	WT	No
DIVD663EVME	pCWR28	0.98	0.46	WT	No
D663A	pCWR29	0.95	0.16	WT	No
K49A D663A	pCWR30	1.07	1.21 *	WT *	No

Fig 5. The RqlH PRTase domain is required for SOS induction in the absence of RqlI and RqlH-associated helicase activity. On the left are schematic representations of the domains within the F11 RqlH protein and mutant variants that were expressed in MG1655 cells. Table summarizes the effects of each RqlH expression construct on bacterial growth (OD_{600}), viability (CFU/ml), cell length, and SOS induction (determined using the reporter strain MG1655 *attTn7*::P_{sulA}-GFP). Results are from 6 h shaking cultures in modified M9 media, and are normalized to control bacteria carrying the empty vector pBAD33. #; phenotype is rescued by co-expression of RqlI. *; phenotype is exacerbated by co-expression of RqlI. See S5 and S6 Figs for full data.

doi:10.1371/journal.ppat.1005317.g005

of filamentous bacteria and the induction of the SOS response, as determined by microscopy and by use of the SOS reporter strain MG1655 *attTn7*::P_{sulA}-GFP (Figs 5 and S6). As seen with F11, the toxic effects of RqlH expression in MG1655 were negated by co-expression of RqlI.

Deletion of the C-terminal PRTase domain (amino acids 509–690) of RqlH abrogated the growth-inhibitory and SOS-inducing effects of RqlH expression in the absence of RqlI (Figs 5, S5 and S6). To examine the role of the PRTase domain more specifically, residues within the putative phosphoribosylpyrophosphate binding pocket of RqlH were mutated. Two PRTase domain variants were tested: one having four conservative amino acid substitutions (DIVD663EVME; [24]) within the putative binding pocket and one having a single alanine replacement (D663A). These residues were chosen based on alignments of RqlH with two phosphoribosyltransferases, PurF and Hpt, and by overlaying the crystal structure of PurF with the predicted structure of RqlH (S7 Fig). Expression of either the DIVD663EVME or D663A RqlH mutant variants had no notable effects on bacterial growth, activation of the SOS response, or filamentation, and had only modest inhibitory effects on the numbers of viable bacteria recovered from the cultures (Figs 5, S5 and S6). These data indicate that the PRTase domain is necessary for RqlH-mediated toxicity, and therefore the defects associated with the Δ*rqlI* mutant. However, expression of the PRTase domain alone was not sufficient to inhibit bacterial growth and did not trigger activation of the SOS response or filamentation, suggesting that additional elements within RqlH cooperate with the PRTase domain to promote toxicity in the absence of RqlI (Fig 5).

To better understand the role of the helicase domain of RqlH, we tested two mutant proteins with single amino acid substitutions (K49A and D148A) that were previously shown to abrogate RqlH helicase activity in *M. smegmatis* [21]. When these mutant proteins were expressed in MG1655, the bacteria grew to a normal density, but low CFU counts were recovered from the cultures (Figs 5 and S5). This reduction in CFUs coincided with increased activation of the SOS response and the development of more filamentous cells by the K49A and D148A mutants (Figs 5 and S6). The loss of viable CFUs was not rescued by co-expression of RqlI, though RqlI did diminish the effects of the K49A and D148A mutations on activation of the SOS response and filamentation. Titers were restored to WT levels if RqlH carried both the K49A mutation

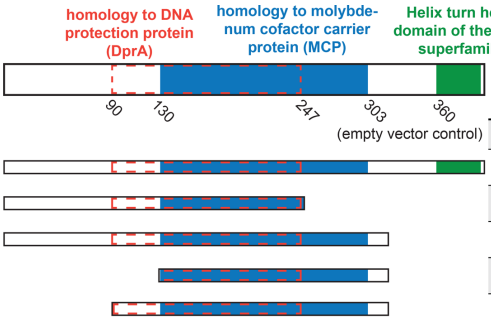
within the helicase domain and the D663A mutation within the PRTase domain (Figs 5 and S5). These data suggest that both RqlI and the helicase domain of RqlH are required to keep the PRTase domain of RqlH in check. However, it is clear that these regulatory interactions are complex. As a case in point, the combined expression of RqlI with the double mutant RqlH K49A/D663A protein leads to a reduction in bacterial titers without a coordinate decrease in culture density, as seen with the single K49A and D148A RqlH single mutants. (Figs 5 and S5). These data are further complicated by the fact that mutations within helicases often have dominant-negative effects that disrupt DNA metabolism [25].

The discordance between culture densities and bacterial CFUs observed with recombinant strains that express the K49A or D148A RqlH mutants ± RqlI, or with bacteria that express both RqlI and the RqlH K49A/D663A double mutant (S5 Fig), is likely in part attributable to stress-induced inhibition of septation and the formation of filamentous cells. Filamentous bacteria will increase the optical density of a culture, but may register as fewer CFUs in plating assays. In our experiments, recombinant strains that reached WT culture densities but had much less than WT titers tended to have more filamentous bacteria (Figs 5, S5 and S6). For the K49A and D148A mutants, filamentation in the absence of RqlI correlated with activation of the SOS response, but this phenomenon was notably less pronounced when RqlI was also expressed (S6 Fig). The co-expression of RqlI with the K49A/D663A RqlH mutant protein had little effect on SOS activation, even though filamentation levels were elevated and CFU counts were markedly reduced. These data indicate that SOS-dependent and SOS-independent filamentation may contribute to the reduced CFU counts recovered following growth of strains that express RqlH helicase domain mutants.

The discrepancies between cell culture densities and titers seen with some of the recombinant strains could also be explained by the generation of anucleate cells or by plating deficiencies, whereby cells may grow well in broth culture but poorly when spread onto agar. To test the former possibility, bacteria were stained using the fluorescent DNA dye Hoechst and imaged. By microscopy, no cells that lack DNA were detected among any of the recombinant strains examined (see S8 Fig as an example). To test for plating deficiencies, bacteria were diluted into fresh LB broth culture at the end of a 6 h growth period and the OD_{600} was tracked over time. Those strains that had decreased CFU counts at the end of the initial 6 h incubation also had a lag in growth when sub-cultured 1:100 into fresh broth, as indicated in S5 Fig as an increase in the time required to reach an OD_{600} of 0.4. These results suggest that the recombinant strains that reach WT density levels in culture, but have reduced titers on agar, have fewer viable cells and have no inherent plating deficiencies.

Overlapping domains within RqlI counter the toxicity of RqlH

As already noted, RqlI contains three putative domains, including a region of homology to DprA, a MoCo binding domain, and a helix-turn-helix (HTH) domain (Fig 1A). To test whether these various domains are required for RqlI function, plasmids encoding different truncation variants of RqlI were created and transformed into F11Δ*rqlI*. Western blots indicated that the recombinant RqlI proteins were expressed, though at varying levels (S9 Fig). As shown in Fig 1B, the growth defect observed with F11Δ*rqlI* under microaerobic conditions can be partially rescued by expression of the full-length RqlI protein (1–400; pCWR16 in Figs 6 and S10A). Expression of the N-terminus of RqlI plus the DprA homology region (1–250; pCWR6) also partially rescued growth of F11Δ*rqlI*. Of note, the 1–250 RqlI mutant complemented F11Δ*rqlI* much like the full-length RqlI even though relatively low levels of the truncated protein were detected (S9 Fig). Nearly complete rescue of F11Δ*rqlI* was attained by expression of an RqlI truncation mutant (1–320) that lacked the C-terminal 80 residues,



Protein	Plasmid	Microaerobic Rescue	RqlH Inhibition
None	pRR48	No	No
1-400	pCWR16	Yes	Yes
1-250	pCWR6	Yes	No
1-320	pCWR7	Yes	Yes
130-320	pCWR11	No	No
85-320	pCWR22	No	No

Fig 6. Domains within the N-terminus of RqlI counter the toxicity of RqlH, independent of the putative HTH domain. The schematics on the left indicate the locations of predicted domains and homology regions within RqlI and various truncation mutants. The table summarizes the effects of each RqlI expression construct on F11Δ*rqlI* titers recovered from 24 h cultures grown microaerobically. Whether or not expression of a given RqlI variant significantly increased bacterial titers in comparison to the empty vector control is indicated under "Microaerobic Rescue". The "RqlH Inhibition" column indicates if a given RqlI variant could counter the toxic effects of RqlH when expressed in MG1655 grown for 6 h under aerobic conditions. See S10 Fig for full datasets.

doi:10.1371/journal.ppat.1005317.g006

including the putative HTH domain (Figs 6 and S10A). These results indicate that the HTH domain is not needed for full RqlI activity under microaerobic conditions, and that the HTH domain may instead actually be somewhat inhibitory to bacterial growth in these assays. In contrast, the expression of RqlI deletion mutants that lack the HTH domain as well as the N-terminal 129 or 84 residues adjacent to the MoCo binding and DprA homology domains did not complement F11Δ*rqlI*.

We also examined the ability of the RqlI variants to abrogate the inhibitory effects of RqlH expression on growth of MG1655 under aerobic conditions (Figs 6 and S10B). Paralleling results obtained with F11Δ*rqlI*, the expression of RqlI truncation mutants that lack the HTH domain as well as the N-terminal 84 or 129 amino acids were unable to counter RqlH, while the production of full-length RqlI or the 1–320 truncation mutant effectively rescued growth of RqlH-expressing MG1655. In contrast, the 1–250 RqlI truncation mutant that restored growth of the F11Δ*rqlI* strain under microaerobic conditions was unable to block RqlH-mediated toxicity in MG1655 under aerobic conditions. In total, these results indicate that the HTH domain is dispensable to the ability of RqlI to interfere with the toxic effects of RqlH under either aerobic or microaerobic conditions. It is also clear that the N-terminus, including both the DprA and MoCo binding domains, is required for full RqlI activity in these assays. However, the C-terminal portion of the putative MoCo binding domain is differentially required, potentially dependent on strain background and/or oxygen levels.

The RqlH and RqlI proteins interact

To test whether the ability of RqlI to keep RqlH activity in check involves physical interaction between the two proteins, co-immunoprecipitations were performed using MG1655 expressing FLAG-tagged RqlI and HA-tagged variants of either full-length RqlH or a C-terminal truncation mutant (1–509) that lacks the PRTase domain (Fig 7). As expected, the RqlH and RqlI proteins were not detected in the control strain carrying the empty plasmid constructs. Full-length RqlH was co-immunoprecipitated with RqlI, and vice-versa, suggesting that these two proteins do associate. This interaction was abrogated in bacteria that express the RqlH 1–509 truncation mutant, indicating that RqlI likely binds the PRTase domain of RqlH. As a control

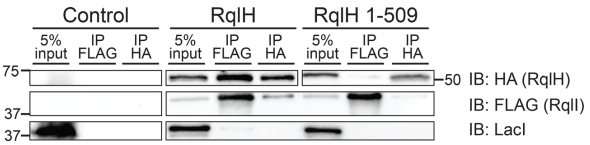


Fig 7. RqlH and RqlI proteins physically interact. MG1655 bacteria that express full-length FLAG-tagged RqlI and HA-tagged RqlH (either full-length RqlH or the 1–509 truncation mutant lacking the PRTase domain) were grown for 4 h, then an additional hour in the presence of IPTG and arabinose, prior to lysis under non-denaturing conditions. As a control, MG1655 carrying the empty vectors pRR48 and pBAD33 were processed in parallel. Lysates were immunoprecipitated using anti-FLAG or anti-HA antibodies, and immunoprecipitates were resolved by SDS-PAGE in preparation for western blot analysis. Aliquots (5%) of the cell lysates that were used as input for the immunoprecipitations were also resolved. Immunoblots (IB) were probed using antibodies specific for the FLAG or HA epitopes or for LacI, as indicated.

doi:10.1371/journal.ppat.1005317.g007

for the specificity of the co-immunoprecipitation procedure, we note that the *lac* operon repressor protein LacI, which is not expected to interact with either RqlH or RqlI, was not pulled down in these assays.

RqlI is required for ExPEC persistence within the gut

Our finding that RqlI is an especially important regulator of RqlH toxicity under low oxygen conditions (Fig 3D) fits well with previous work in which we showed that RqlI is critical to ExPEC survival within the urinary tract and bloodstream [14]. These niches, like many sites of infection, have limiting amounts of oxygen available for use by facultative anaerobes like ExPEC. Since a primary reservoir of ExPEC within vertebrate hosts is the intestinal tract, in which oxygen levels are generally low [26], we set out to determine if RqlI was also important to the fitness of F11 within the gut. For these studies, we used marked strains that carried chromosomal Clm^R or Kan^R cassettes. Adult Balb/c mice were inoculated via oral gavage with 10^9 CFU of F11 or F11Δ*rqlI*, and survival was tracked by titration of fecal pellets on selective plates. WT F11 persisted within the gut at fairly steady levels (medians ~10^6 CFU/g feces) for more than 10 d, while titers of the Δ*rqlI* mutant were significantly reduced as early as day 1 (Fig 8A). By the end of the experiment, only 2 of the 10 mice that received F11Δ*rqlI* remained colonized with the mutant. The persistence defect observed with F11Δ*rqlI* was also evident when the mutant was co-inoculated 1:1 with WT F11. In these competitive assays, the WT strain outnumbered the Δ*rqlI* mutant by more than a 1,000-fold within 3 d (Fig 8B).

To test if RqlI is important for gut colonization by other ExPEC strains, *rqlI* was deleted from the urosepsis isolate CFT073 and from the pyelonephritis isolate 536. In competition assays, CFT073Δ*rqlI* titers within the intestinal tract were greatly diminished within 3 d of inoculation, mirroring the situation seen with F11 (S11A Fig). While 536Δ*rqlI* was also defective in gut colonization, the decline of this mutant was more gradual and less pronounced, decreasing about a 100-fold relative to the WT 536 strain by day 7 post-inoculation (S11B Fig). These results, coupled with our previous findings [14], indicate that RqlI promotes the colonization and persistence of various ExPEC strains within distinct host environments, including the urinary tract, the bloodstream, and the intestinal tract.

To determine if the *in vivo* defects observed with the Δ*rqlI* mutants are attributable to the disregulation of RqlH activities, as seen in our *in vitro* assays, F11 with a complete deletion of the *rqlHI* operon was competed 1:1 against the WT F11 strain in our mouse models for gut colonization, UTI, and bacteremia. Following oral inoculation of Balb/c mice with a 1:1 mixture of F11 and F11Δ*rqlHI*, fecal bacterial titers were tracked over the course of 50 days (Fig 8C).

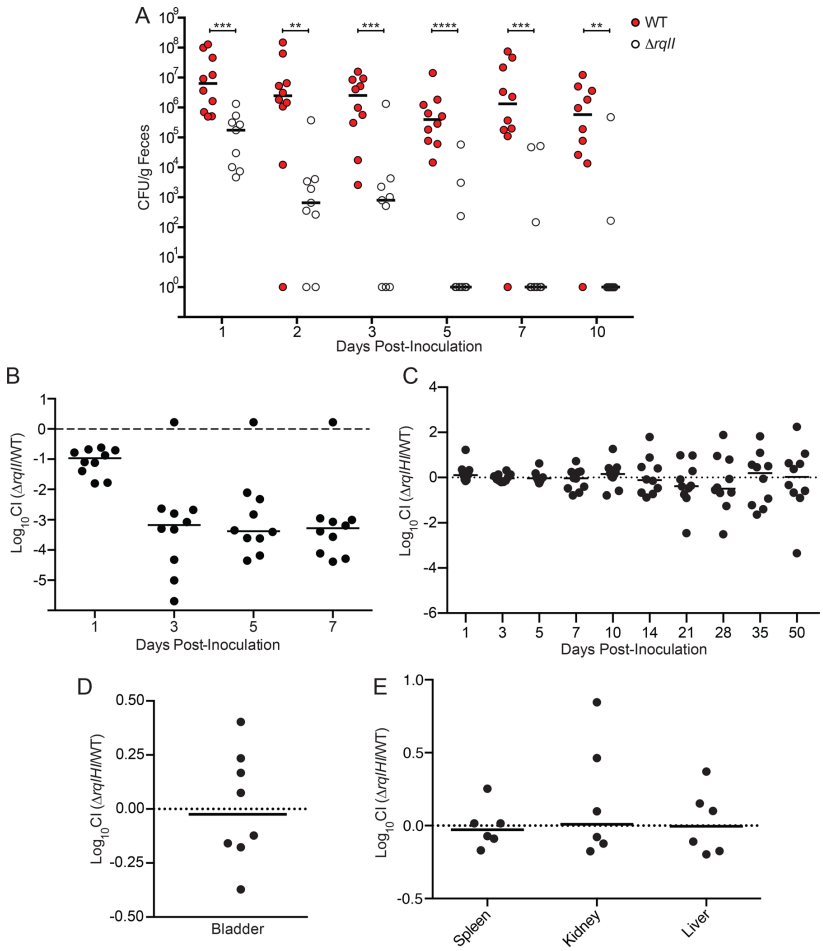


Fig 8. Decreased fitness of Δ*rqlI* mutants within diverse host niches is attributable to RqlH. (A) WT F11 or F11Δ*rqlI* were inoculated into Balb/c mice via oral gavage and feces were collected, homogenized, and plated to determine bacterial titers at the indicated time points. For competitive assays, (B) F11 and F11Δ*rqlI* or (C) F11 and F11Δ*rqlHI* were mixed 1:1 prior to delivery into mice via oral gavage. Graphs show competitive indices (CI), as determined by dilution plating of fecal homogenates on selective media at the indicated time points. (D) F11 and F11Δ*rqlHI* were inoculated transurethrally as a 1:1 mixture into the bladders of Balb/c. CI values were calculated from bacterial titers recovered from bladders at 3 d post-inoculation. (E) F11 and F11Δ*rqlHI* were mixed 1:1 and injected intraperitoneally into mice to initiate systemic infections. At 12 h post-inoculation, the spleen, liver, and a kidney were isolated from each mouse, homogenized, and titered to determine the CI of the Δ*rqlHI* mutant. Bars indicate median values.

doi:10.1371/journal.ppat.1005317.g008

Median competitive indices stayed close to zero throughout the experiment, indicating that the Δ*rqlHI* mutant does not have a defect in mouse gut colonization, in sharp contrast to the single

Δ*rqlI* mutant. Defects previously observed with F11Δ*rqlI* survival within the bladder and the bloodstream of mice [14] were also ablated if *rqlH* was deleted together with *rqlI* (Fig 8D and 8E). In total, these data indicate that the toxicity associated with RqlH expression in the absence of RqlI in our *in vitro* assays is also manifest *in vivo* within diverse host environments. Although the *in vivo* defects associated with the *rqlI* mutant are not particularly surprising in light of our *in vitro* findings, it is noteworthy that the *in vivo* effects are generally greater in magnitude than those seen *in vitro*.

RqlH and RqlI orthologs are found across multiple genera, but are not always linked

Given the close relationship between *rqlH* and *rqlI* that we observed in F11, we wondered if these proteins are always present as a pair within bacterial genomes, if they are always adjacent to one another, and if they are widely distributed among bacteria. A search for orthologs of RqlH and RqlI was undertaken using NCBI sequence databases, followed by analysis of the results using custom Python scripts. We defined an ortholog as a hit that aligns with at least 75% of the RqlI or RqlH protein sequence and that has a percent positive score of at least 50 and an E-value of 1e-06 or better. In addition, putative orthologs had to successfully return the original query protein (RqlI or RqlH) as the top hit in reciprocal BLAST searches against F11. For a list of RqlI and RqlH orthologs, and for the raw data used in our analysis, see S1 Dataset.

Our results indicate that RqlH or RqlI orthologs are distributed across a phylogenetically diverse range of bacteria, with a total of 240 distinct genera represented in the analysis. Unexpectedly, the majority of bacterial strains (55.3%) encoded an ortholog of RqlH, but had no detectable ortholog of RqlI (Fig 9A). There was also a small portion (1.2%) of the bacteria that contained only RqlI. 43.5% of bacteria within the database encoded both RqlI and RqlH.

By analyzing different bacterial subsets, we found that there is a striking difference in the distribution of RqlH and RqlI orthologs in Gram-positive versus Gram-negative. The majority (87.9%) of Gram-negative bacteria encode both RqlI and RqlH, with just 9.7% carrying only RqlH (Fig 9B). In contrast, nearly all (99.6%) Gram-positive bacteria encode only RqlH, while the remaining few have both proteins (Fig 9C). Bacteria with only an RqlI ortholog were found exclusively among the Gram-negative bacteria (Fig 9B). These observations were echoed in our analysis of specific individual genera. For example, 98.6% of the *Escherichia* strains (Gram-negative) encode both RqlH and RqlI (Fig 9D), while very few have only RqlH or only RqlI. *Streptomyces* species—like most other Gram-positive taxa—encode only RqlH, with no detectable orthologs of RqlI in any of the strains (Fig 9E). This is also the case for all strains of *Mycobacterium smegmatis*, the Gram-positive acid-fast species in which RqlH was first identified [21].

Within strains that have both *rqlH* and *rqlI* orthologs, we found that the genes are always oriented with *rqlH* upstream of *rqlI*, reminiscent of the operon structure in F11. In most of these strains, the start of *rqlI* is close to or overlapping with the terminus of *rqlH*, though in some strains the two genes are separated by more than 2 kbp. Fig 9F shows examples of the variable arrangements of *rqlH* and *rqlI* within divergent species. In *Klebsiella oxytoca* M5aI and *Pseudomonas aeruginosa* VRFPA06, the *rqlH* and *rqlI* genes are closely associated, or even slightly overlapping, whereas within the cyanobacterium *Trichodesmium erythraeum* IMS101 the two genes are separated by an unannotated 2196 bp sequence. Occasionally, we identified other annotated elements inserted between the *rqlH* and *rqlI* loci, as in *Shigella sonnei* 53G where sequences encoding the PRTase domain of RqlH have been separated from the helicase domain by an insertion element and a putative transposase. Together, these observations demonstrate that *rqlH* and *rqlI* orthologs, when found together within the same genome, are largely syntenic with the F11 *rqlHI* operon. The exceptions to this arrangement, as in *Shigella sonnei*

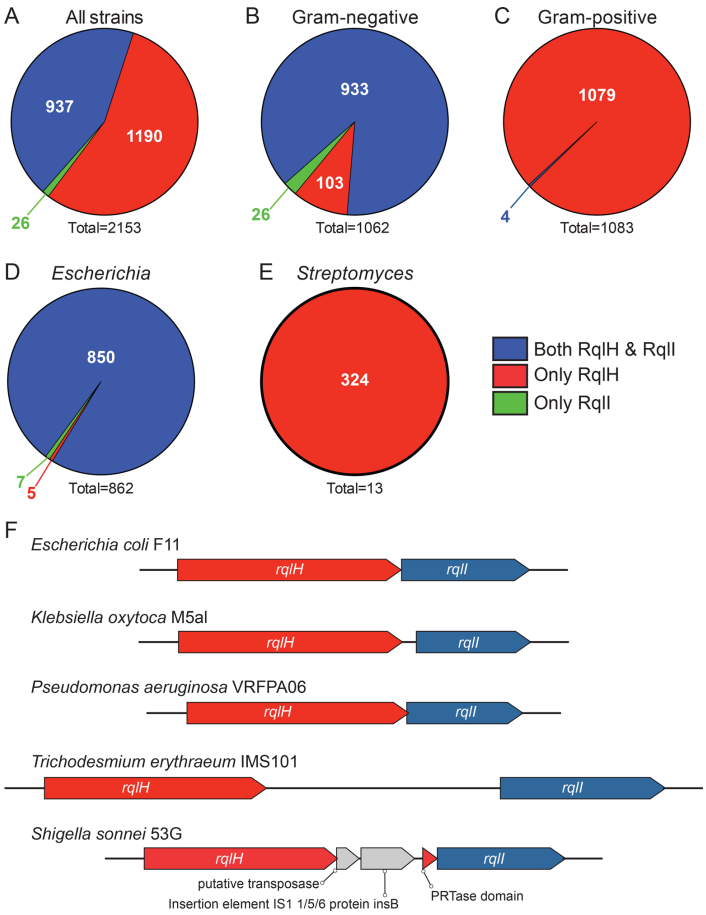


Fig 9. RqlH and RqlI homologs are found across various genera, either alone or together. F11 RqlH and RqlI homologs were found using blastp and tblastn searches of various NCBI sequence databases. (A-E) Distribution of strains containing only a homolog of RqlI (green), only RqlH (red), or both RqlH and RqlI (blue). The numbers of bacterial strains that are associated with each pie section are indicated, with totals given beneath each graph. Included are (A) distributions of RqlH and RqlI orthologs in all strains, (B) among Gram-negative bacteria, (C) among Gram-positive organisms, (D) within the *Escherichia* genus, or (E) as part of the *Streptomyces* genus. (F) Examples showing the arrangements of *rqlH* and *rqlI* orthologs in different bacterial species.

doi:10.1371/journal.ppat.1005317.g009

53G, may reflect alternate ways in which bacteria have evolved to modulate the potentially toxic effects of RqlH expression.

Discussion

This study was aimed at functionally defining *EcF11_3933*, a hypothetical gene that we previously identified as an important facilitator of ExPEC fitness in zebrafish infection models and in mouse models of UTI and sepsis [14]. Homologs of this gene are also expressed by UPEC isolates recovered directly from the urine of women with UTI [27]. Results presented here indicate that the *EcF11_3933* gene product, which we have renamed RqlI for RecQ-Like Helicase Interactor, is able to bind to and modulate the activity of the RecQ-Like Helicase RqlH. We propose a model in which RqlI and RqlH act cooperatively to perform an as-yet undefined beneficial function on bacterial DNA (Fig 10). Although more work is required to further vet this idea, the hypothesis is supported by several pieces of data. First, the RqlH homolog in *M. smegmatis* is known to be a helicase belonging to the RecQ family [21], which includes several helicases that perform general maintenance functions on prokaryotic and eukaryotic genomes [28]. Second, RqlH and RqlI physically interact (Fig 7) and are encoded within the same operon (Fig 4A and 4B). Third, the Δ*rqlI* mutant experiences several defects that are attributable to RqlH activities, all of which involve phenotypes associated with DNA stress. These include induction of the SOS response, increases in cell length (filamentation), higher sensitivity to DNA damaging agents, and elevated mutation rates (Figs 2 and 3). We suggest that in the presence of a partially functional or disregulated RqlH-RqlI system, some harmful DNA byproduct(s) is created that leads to a loss in cellular growth and viability (Fig 10B). Interestingly, the defects associated with disruption of the RqlH-RqlI system are exasperated under low oxygen conditions, suggesting possible links with anaerobic respiration and/or other redox-sensitive processes.

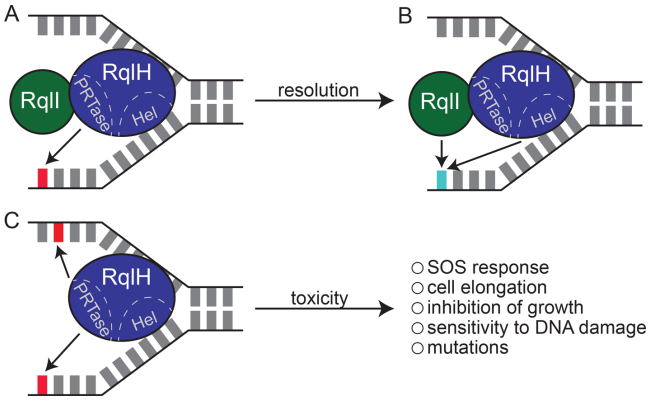


Fig 10. In the absence of functional RqlI the PRTase domain of RqlH elicits genotoxic effects. Model depicts a possible function for the *rqlHI* operon. (A) The PRTase domain of RqlH alters a nucleotide (red) within the bacterial genome, and (B) RqlI functions to resolve this modified nucleotide (light blue) with the help of the helicase domain of RqlH. The resolution event could include further modification of the nucleotide or restoring the nucleotide to its original structure. (C) In the absence of RqlI, the PRTase domain of RqlH continues to alter chromosomal nucleotides, but these modifications are not resolved, producing effects that are toxic to the bacteria, resulting in the induction of the SOS response, cell elongation, growth inhibition, heightened sensitivity to DNA damaging agents, and increased mutation rates. Many of these phenotypes are intensified under low oxygen conditions for unknown reasons.

doi:10.1371/journal.ppat.1005317.g010

Our observations and the model presented in Fig 10 raise a number of questions concerning the RqlH-RqlI system. Chief among these is the question of function. We were unable to detect any decrease in bacterial fitness in Δ*rqlHI* mutants in our *in vitro* experiments or in mouse models of infection, using either competitive or non-competitive assays (Figs 8 and S11). These results may reflect limitations in the resolution of the mouse models, or their inability to recapitulate all aspects of natural colonization and infection processes. For example, our assays employed human ExPEC isolates that may use different strategies to colonize the human versus the murine intestinal tract. The intestinal microbiota associated with humans and mice are often similar at the phyla level, but can be quite different qualitatively and quantitatively at lower taxonomic levels [29,30]. Such differences may affect the types of nutrients and stresses encountered by ExPEC within the gut and could potentially influence the requirements for genes like *rqlHI*. ExPEC are versatile organisms, and so it is also feasible that RqlH-RqlI are important to the survival of these pathogens within environmental reservoirs or at other sites that were not modeled in our study.

The key to understanding the role of the RqlH-RqlI system in bacterial physiology may be the PRTase domain of RqlH. Indeed, RqlH is unique among helicases because it possesses a PRTase domain in addition to its helicase domain. PRTase domains can be found in proteins that function in nucleotide salvage pathways by adding phosphoribosyl groups to spent nucleotides. As an example, the guanine-xanthine phosphoribosyltransferase (Gpt) of *E. coli* catalyzes the addition of phosphoribosyl pyrophosphate (PRPP) to guanine to make the nucleotide guanosine monophosphate (GMP) [31]. Given the combination of a helicase with a PRTase domain in a single protein, it is tempting to speculate that RqlH functions to add PRPP directly to DNA, or that it adds PRPP to nucleobases while in association with DNA.

The function of *rqlHI* may be related to uptake of foreign DNA, as there are at least two connections between the RqlH-RqlI system and natural transformation in competent bacteria. First, RqlI contains a region of homology to DprA, a protein that has been shown to protect ssDNA during the translocation of foreign DNA into naturally competent bacteria [32]. Of note, in addition to the DrpA homology region within RqlI, F11 and other ExPEC isolates encode a DprA protein that is more closely related to the canonical DprA. However, DprA has no detectable effect on transformation in a K-12 *E. coli* strain [33]. The second connection to natural transformation is the PRTase domain of RqlH, a domain that is found in a number of proteins involved in competence in other bacteria. In *Bacillus subtilis*, two proteins encoded by the *comF* operon are suggestive of the functional domains contained within RqlH. Specifically, the *comF* operon, which *B. subtilis* requires for competence, codes for a putative helicase and a predicted PRTase separated by an intervening gene with no known domains [34]. The Com cluster of genes that facilitate transformation in *Haemophilus influenzae* also includes a gene (ComF) that is homologous to the PRTase domain of RqlH, but it is seemingly not associated with any nearby helicase [35]. Arguing against a role of the *rqlHI* operon in natural transformation is the fact that it is generally thought that *E. coli* is not naturally competent, though it can be made chemically competent or electrocompetent in a laboratory setting. Nonetheless, incidences of natural transformation by *E. coli* have been reported [36,37], but we have been unable to detect any transformation events using our ExPEC isolates, in agreement with a recent report [38].

In light of our results showing that RqlH is toxic to *E. coli* strains in the absence of RqlI (Fig 3), it was surprising to find that many bacteria carry RqlH but do not encode RqlI homologs (Fig 9). In particular, nearly all Gram-positive organisms that encode an RqlH homolog lack any identifiable RqlI ortholog (Fig 9C). These observations may be explained in several ways. It could be that RqlH on its own is not inherently toxic to some bacteria (e.g. Gram-positive microbes) due to differences in their physiology in comparison with ExPEC and the K12 strain MG1655. It is also possible that bacteria that lack *rqlI* express different, less toxic *rqlH* alleles

than found in bacteria that encode both genes. However, we did not detect major differences in the sequences of RqlH proteins from bacteria that have just RqlH versus those with both RqlH and RqlI. Finally, RqlH activity could be kept in check by other proteins that are more distantly related to RqlI. In this same vein, it is interesting to note that RqlI orthologs are more prevalent among known pathogens (83.3% of RqlI homologs are encoded by pathogens in the TEA proteomic database; see [14]), while RqlH orthologs are much less often pathogen-associated (39.1%). Altogether, these observations suggest that RqlH function could vary depending on bacterial strain background and the presence or absence of RqlI.

The observation that RqlH expression in the absence of RqlI is toxic to *E. coli* strains is reminiscent of a type II toxin-antitoxin (TA) system. These TA systems consist of a protein toxin that is more stable than its cognate antitoxin protein, which binds to and inhibits the toxin [39]. TA systems were originally found encoded on plasmids where they promoted plasmid maintenance. In any cell that spontaneously loses the plasmid, the TA system is no longer transcribed, and the remaining toxin protein outlasts the residual antitoxin, resulting in bacterial cell stasis or death. In this and other similar situations, TA systems can be thought of as selfish genetic elements that function only to proliferate without providing any benefit to the host bacterium. More recently, TA systems have also been shown to have beneficial functions that can protect bacteria from stressors. For example, TA systems can promote the formation of antibiotic-insensitive persister cells [39] and select TA systems in ExPEC can promote stress resistance and colonization of the urinary tract [40].

Results presented here indicate that RqlH and RqlI can in many ways act like a TA system, meshing with data from others showing that an RqlH ortholog (PsyrT) in *Pseudomonas syringae* is toxic to the bacteria in the absence of the RqlI ortholog PsyrA [41]. However, we note that if RqlH and RqlI are to be defined as a TA system, it would be non-canonical for the following reasons: 1) RqlH and RqlI are much larger (698 and 400 amino acids, respectively) than typical TA system proteins (~100 amino acids); 2) *rqlH* is situated upstream of *rqlI*, whereas in most type II TA systems the antitoxin is encoded upstream of the toxin; and 3) an Δ*rqlHI* mutant exhibits increased survival (persisters) in the face of antibiotic treatments (S12 Fig), in contrast to the results obtained when *bona fide* TA systems are deleted [40]. Finally, we suggest that the distribution among most bacteria of RqlH orthologs that are encoded in the absence of RqlI argues against the possibility that the RqlH-RqlI system always functions as a TA-like genetic element, although it can resemble one within ExPEC strains.

Cumulatively, this work illustrates the complexities of assigning function to hypothetical genes that are identified as fitness determinants in high-throughput screens, such as Tn-seq. The ability of RqlI to counter the oxygen-sensitive genotoxic effects associated with RqlH helps explain why ExPEC strains are dependent on RqlI expression within the microaerobic confines of the gut and other extraintestinal niches. Specifically, our work indicates that RqlI functions to modulate the context-dependent toxicity of the RqlH helicase and its PRTase domain, but opens up many more questions concerning the functional relevance of the RqlH-RqlI system among ExPEC strains and of the numerous other RqlH and RqlI orthologs identified among thousands of phylogenetically diverse species. Future analyses detailing the regulation and functionality of the RqlH-RqlI system and its many homologs may provide the basis for novel anti-bacterial therapeutics that can unleash the toxic potential of RqlH.

Materials and Methods

Ethics statement

Mice were handled in accordance with protocols approved by the Institutional Animal Care and Use Committee at the University of Utah (Protocol number 10–02014), following US

federal guidelines indicated by the Office of Laboratory Animal Welfare (OLAW) and described in the Guide for the Care and Use of Laboratory Animals, 8th Edition.

Bacterial strains, media, and plasmids

The ExPEC strains used in this study included the cystitis isolate F11, the urosepsis isolate CFT073 and the pyelonephritis isolate 536, as well as the K-12 strain MG1655 (S1 Table). Manipulations of chromosomal DNA, including both the generation of knockout strains and the knockin strain MG1655 *attTn7*::P_{sulA}-GFP were achieved with lambda-Red-mediated recombination using plasmid pKM208 [42]. Knockout strains were produced using PCR products containing ~40 bp overhang regions with homology to target loci. To create the MG1655 *attTn7*::P_{sulA}-GFP reporter strain, a *tetA-sacB* cassette was amplified from T-SACK cells [43], and combined via PCR with two 500 bp arms homologous to the *attTn7* site downstream of *glmS*. Next, the *tetA-sacB* cassette was replaced with the P_{sulA}-GFP reporter from pJLJ3 with selection on Tet/SacB counter-selection agar, as described [43]. Primers used to create knockout and knockin constructs, as well as those used for mutant confirmation, are listed in S1 Table. All PCR products were purified with the DNA Clean & Concentrator-5 kit (Zymo Research, catalog #D4004). Where indicated, antibiotic resistance cassettes were removed using the flippase-expressing pCP20 plasmid, as described [44].

Bacteria used for genetic manipulations were grown in LB broth, whereas bacteria used for both the *in vivo* and *in vitro* experiments were grown in modified M9 media. The modified M9 media contained M9 salts (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, and 0.5 g/L NaCl), as well as 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1% glucose, 0.00125% nicotinic acid, 0.00165% thiamine, and 0.2% casein amino acids.

All plasmids used in this study are listed in S2 Table. For previously unpublished plasmids, the primers utilized in their creation are also indicated. RqlI expression vectors were created by inserting PCR products into the PstI and HindIII sites of pRR48, under control of an IPTG-inducible promoter [45]. RqlH expression vectors were made by insertion at the SacI and HindIII sites within pBAD33, downstream of an arabinose-inducible promoter [46]. Plasmids were generated using standard cloning techniques, the QuikChange II XL Site-Directed Mutagenesis Kit by Agilent Technologies, or by overlap extension PCR [47], as indicated in S2 Table.

Colonization of mice with ExPEC

Female Balb/c mice were purchased from Jackson Labs and used between 7–8 weeks of age, and all experiments were performed in accordance with IACUC-approved protocols. All mouse experiments were repeated at least twice, and the total combined data from six or more animals is presented. Bacteria used to inoculate mice were grown statically at 37°C in 20 mL modified M9 in 250 mL Erlenmeyer flasks for 24 h, pelleted by spinning at 8000 r.c.f. for 8 min, washed, and resuspended in phosphate buffered saline (PBS). For competitive assays, WT and mutant bacteria were mixed in equal parts prior to inoculation. These experiments used bacteria carrying either a chloramphenicol or kanamycin resistance cassette, allowing them to be easily selected and distinguished from other bacteria within the microbiota, and enabling the facile tracking of individual WT and mutant strains in competitive experiments.

To assess intestinal colonization by the ExPEC strains and their mutant derivatives, mice were orally gavaged with 50 μl of a bacterial suspension containing $1x10^9$ CFU of bacteria in PBS. Fecal samples were collected at various time points post-gavage, homogenized in 1 mL 0.7% NaCl, and briefly spun to pellet any insoluble debris. Supernatants were then serially diluted and plated on LB agar containing chloramphenicol (10 μg/ml) or kanamycin (50 μg/ml). Fecal

samples were also collected and plated prior to gavage to ensure that the resident microbiota did not include any culturable chloramphenicol- or kanamycin-resistant bacteria.

For the UTI model, mice were anesthetized using isofluorane inhalation and slowly inoculated via transurethral catheterization with 50 μl of a bacterial suspension containing ~$1x10^8$ bacteria in PBS. After 3 d, the bladders were extracted and homogenized in PBS. Homogenates were then serially diluted and plated.

To initiate bacteremia/sepsis, mice were injected intraperitoneally with 200 μl of PBS containing ~$1x10^7$ bacteria. The animals were monitored for signs of extreme morbidity and euthanized at 12 h post-inoculation. The livers, kidneys, and spleens were then recovered and homogenized in PBS. Bacterial titers present within the various tissues were quantified by plating serial dilutions of the homogenates.

Bacterial growth and reporter assays

To quantify bacterial growth in various oxygen levels, bacteria were grown overnight in M9 in loose-capped tubes, then subcultured the following day 1:100 in M9. During the subculture stage, the bacteria were grown on a shaking incubator for 24 h in 12-well plates either in aerobic conditions or in a Mitsubishi AnaeroPack 2.5 L Rectangular Jar holding a AnaeroPack-Microaero sachet or AnaeroPack-Anaero sachet to model microaerobic and anaerobic conditions, respectively. To calculate generation time, the bacteria were titered at two time points during exponential phase.

To measure *sulA*-GFP expression in F11 bacteria carrying the pJLJ3 plasmid or the pJLJ1 empty vector control, bacteria were grown overnight in loose-capped tubes in M9 with kanamycin, then subcultured 1:100 into fresh M9 with kanamycin. Cells were then grown microaerobically and anaerobically for 24 h in 12-well plates, or aerobically for 4 h in loose-capped tubes. For the control conditions, mitomycin C was added to a final concentration of 0.25 μg/mL in aerobic cultures at 3 h, and then incubated for another hour. Both GFP fluorescence and the OD_{600} were measured using a BioTek Synergy H1 plate reader.

For experiments with the MG1655*attTn7*::P_{sulA}-GFP strain carrying various RqlH mutant proteins, the cells were grown overnight in M9 with ampicillin and chloramphenicol in loose-capped tubes The bacteria were then subcultured 1:100 into M9 with ampicillin, chloramphenicol, IPTG (1 mM), and arabinose (0.05%), and incubated aerobically in loose-capped tubes for 6 h. A growth curve was performed by again subculturing cells after the 6 h growth period 1:100 into fresh LB. The OD_{600} was then read every 30 minutes by a Bioscreen C machine.

In the assays, 1 mM IPTG or 0.05% arabinose was added to the media to induce expression of RqlI, RqlH or their derivatives.

DNA damage sensitivity assays

Following overnight growth from frozen stocks, F11, F11Δ*rqlH*, F11Δ*rqlI*, and F11Δ*rqlHI* were diluted 1:100 into modified M9 media and grown with shaking (aerobically) for 4 h in loose-capped tubes. After addition of mitomycin C (0.25 μg/mL), the incubations were continued for one more hour. Bacterial titers before and after mitomycin C treatment were determined by plating serial dilutions. Survival rates were calculated as the numbers of viable bacteria present after mitomycin C treatment as a percent of bacteria present prior to addition of the toxin.

To assess strain susceptibility to UV irradiation, bacteria were grown aerobically in loose-capped tubes from overnight cultures for 4 h at 37°C, then serially-diluted and spread onto LB plates in order to obtain ~100 cells per plate. Plates were then exposed to 0, 25, or 50 J/m^2 UV light produced by a Stratalinker 1800. The plates were then incubated overnight and surviving bacteria quantified.

Microscopy

To visualize bacteria, 5 μl of cells (~5x10^7 cells) resuspended in PBS containing 1 μg/mL Hoechst dye were spread on a glass slide and incubated at room temperature until the PBS had evaporated. A drop of FluorSave Reagent was added to each slide and a coverslip was placed on top. Bacteria were imaged by fluorescence or phase-contrast microscopy using an Olympus BX51 microscope equipped with a 100X oil immersion objective and a QImaging QIClick Cooled CCD camera. To measure cell lengths, straight-line measurements from one tip of a cell to the other were made using ImageJ [48]. For each bacterial strain, 6–10 fields and more than 200 bacteria were examined.

Transposon screen

A P_{sulA}-*lacZ* reporter plasmid (pCWR2) was created by overlap extension PCR using a promoterless *lacZ* plasmid (pCWR1) as a template (S2 Table). The reporter plasmid was introduced into F11Δ*rqlI*Δ*lacZY*::*clm*, and the resulting strain was randomly mutagenized by conjugation with EcS17/pSAM-Ec donor bacteria as described [14]. For conjugation, bacteria were grown overnight in LB broth, and 1 mL of donor cells were mixed with 0.5 mL of recipient cells, washed once with LB broth, and then spread onto an LB plate. After a 5 h incubation at 37°C, bacteria were recovered from the plate surface, resuspended in 1 mL of LB broth, and diluted 1:100. Aliquots (100 μl) were plated onto tetrazolium lactose plates containing ampicillin, chloramphenicol, and kanamycin to select for transposon-mutagenized F11Δ*rqlI*Δ*lacZY*::*clm*/pCWR2 bacteria. Colonies that were darker than the F11Δ*rqlI* Δ*lacZY*::*clm*/pCWR2 strain were re-streaked onto new tetrazolium lactose plates to verify decreased P_{sulA}-*lacZ* activity. Transposon mutants that returned *sulA* expression to WT levels underwent a secondary screen in a broth-based modified Miller assay, as described [49]. For mutants that exhibited decreased *sulA* expression on both tetrazolium lactose plates and in the Miller assays, transposon insertion sites were mapped by arbitrary PCR. Briefly, 30 μl colony PCR reactions were carried out using the pSAM-EC kan us1 primer paired with each of the arbitrary primers ARB1A, ARB1B, and ARB1C (S3 Table). The thermocycler program for this reaction was as follows: 95°C for 5 min; 5 cycles of 94°C for 30 sec, 30°C for 30 sec, and 72°C for 1:30; 30 cycles of 94°C for 30 sec, 45°C for 30 sec, and 72°C for 2 min; and ending with a 5 min incubation at 72°C. A 2 μl aliquot of this reaction was then used as template in a second PCR reaction using the pSAM-EC kan us2 and ARB2 primers, in a total volume of 50 μl. The second thermocycler program was as follows: 94°C for 2 min; 30 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 1:30; and then 72°C for 5 min. The entire volume of this second reaction was resolved on an agarose gel, and the largest distinct band was gel-purified and sent for Sanger sequencing using the primer pSAM-EC kan us3.

Fluctuation assays

Bacteria were grown overnight in modified M9 media and then sub-cultured 1:100 into several 1 mL cultures. These were grown for either 4 h aerobically in loose-capped tubes or for 24 h microaerobically in 12-well plates. A few cultures were serially diluted and plated onto LB agar to quantify the total number of cells. The entirety of each remaining culture was then spread onto LB plates containing 20 μg/mL nalidixic acid. The following day, colonies were counted to quantify total bacterial titers as well as the numbers of nalidixic acid resistant cells. Results were analyzed by the online FALCOR calculator for fluctuation assays to determine mutation rates [50].

RT-PCR

To probe for *rqlHI* mRNA, WT F11 was grown aerobically for 4 h in loose-capped tubes in modified M9 media after being diluted 1:100 from an overnight culture. RNA was extracted from cells using a Norgen Total RNA Purification Kit, then treated with DNase for 1.5 h, precipitated, and resuspended in water. To create cDNA, 0.5 μg of RNA was used with the Superscript III First-Strand Synthesis Kit (Life Technologies). Control reactions lacking reverse transcriptase were run concurrently. The cDNA and control reactions were then used as template in PCR reactions using two primer sets that spanned the *rqlH* and *rqlI* gene junction, or with one primer set internal to *rqlI* (S3 Table). For RT-qPCR, RNA was extracted from bacteria grow under aerobic or microaerobic conditions and cDNA produced as described above. A LightCycler 480 instrument was used to run the qPCR reactions using primers specific to *frdA*, *rqlH*, or *rqlI*, or primers that span *rqlH* and *rqlI* (S3 Table). Results were normalized to 16S rRNA levels.

Co-immunoprecipitations and western blots

For analysis of protein complexes, a control MG1655 strain carrying the empty vectors pRR48 and pBAD33, or recombinant MG1655 strains expressing FLAG-tagged RqlI and HA-tagged RqlH or the RqlH (1–509) truncation mutant were diluted 1:100 from overnight cultures into 20 mL minimal M9 media with ampicillin and chloramphenicol. After 4 h of growth at 37°C, IPTG (1 mM) and arabinose (0.05%) were added to induce expression of RqlI and RqlH, respectively. Cultures were grown for 1 h more and bacteria were then pelleted by centrifugation and frozen. After thawing, bacterial pellets were resuspended in 500 μL B-PER Bacterial Protein Extraction Reagent (Life Technologies) containing 1X cOmplete protease inhibitor cocktail (Roche), 1 mM PMSF, and 1 μL/mL Lysonase (Merck Millipore), and then incubated at room temperature for 10 min. A small aliquot of each lysate was set aside for use as input controls for the immunoprecipitations. Lysates were incubated with Dynabeads Protein G (Life Technologies) pre-loaded with mouse anti-HA (Santa Cruz; SC-7392) or rabbit anti-FLAG (Sigma Aldrich; F7425) antibodies. After three washes in PBS, bead-protein complexes were resuspended in 30 μL sample buffer (4% sodium dodecyl sulfate [SDS], 20% glycerol, 0.02% bromophenol blue, 10% 2-mercaptoethanol, 0.125 M Tris-HCl [pH 6.8]) and heated for 5 min at 95°C in preparation for SDS–polyacrylamide gel electrophoresis (PAGE). Input controls (5% of initial lysates) for the co-immunoprecipitation experiments were similarly prepared for SDS-PAGE. Following electrophoresis, proteins were transferred to Immobilon-FL PVDF membranes (Millipore) that were then probed using antibodies specific for the FLAG or HA epitope tags or for LacI (AbCam; ab33832). Blots were developed using HRP-conjugated secondary antibodies with the BM Chemiluminescence Blotting Substrate (Roche) and imaged using a BioRad ChemiDoc MP system.

To analyze expression of the epitope-tagged RqlH and RqlI proteins and their variants, or to confirm GFP expression levels by the reporter strain MG1655*attTn7*::P_{sulA}-GFP, bacterial cells were lysed and protein concentrations were determined as described above. Proteins from each sample were resolved by SDS-PAGE, transferred to PVDF membranes, and probed using anti-FLAG, anti-HA, or anti-GFP (Santa Cruz; SC-9996) antibodies. Subsequently, blots were stripped and probed using anti-*E. coli* antibody (BioDesign) to ensure equal protein loading.

Distribution of RqlH and RqlI proteins

To find homologs of RqlH and RqlI among other bacteria, the F11 sequence of each protein was used as a query for an online blastp search using the non-redundant protein sequences database. Additionally, tblastn searches were performed against the non-redundant nucleotide

and whole genome shotgun databases in order to find unannotated orthologs that would not be present in the non-redundant protein database. All BLAST searches were done with an expect threshold cutoff of 1e-06, and custom Python scripts utilizing the Biopython module [51] were used to parse the results. The hits were reciprocally blasted against the F11 proteome, and those that brought back the original query protein (i.e. RqlH or RqlI) as the top hit, aligned with at least 75% coverage, and had a percent positive score of at least 50, were defined as homologs. The species that carried homologs of only RqlH, only RqlI, or both were enumerated. To examine synteny of the *rqlH* and *rqlI* genes within bacteria that carried both genes, only hits originating from the blastp search were considered. Protein GenBank records were accessed via EFetch with Biopython, and parsed to obtain the accession numbers of the nucleotide sequences containing each *rqlH* or *rqlI* ortholog and its location within its respective genome. The gene locations were then compared to determine orientation and proximity of *rqlH* and *rqlI* to each other.

Persister assays

Persister cell assays were performed as described [40]. Briefly, bacteria were diluted 1:100 from overnight cultures into 5 mL of LB broth and then grown with shaking (225 rpm) at 37°C for 2 h. At this point, an aliquot of each culture was taken for CFU enumeration prior to addition of ampicillin (100 μg/mL) or ciprofloxacin (10 μg/mL). After a further 5 h incubation, 1 mL of each culture was pelleted, washed once with LB broth, and surviving bacteria were determined by plating serial dilutions. Percent survival, reflecting the numbers of persister cells in each sample, was calculated by dividing the number of viable bacteria present after antibiotic treatment by the number present prior to antibiotic addition.

Statistics

Results from *in vivo* competition assays were analyzed by one sample T tests. Results from non-competitive assays in mice were analyzed by Mann-Whitney two-tailed *t* tests. Results from *in vitro* assays were analyzed by unpaired Student's *t* tests. Prism 5.01 (GraphPad Software, Inc.) was used for all statistical tests, including Student's *t* tests. P values of less than 0.05 were defined as significant.

Supporting Information

S1 Table. Strains used in this study. Includes a list of the strains used in this study and primers used to create and verify them, if applicable.
(PDF)

S2 Table. Plasmids used in this study. Includes a list of the plasmids used in this study. If a plasmid was made for this study, the primers and method used to create it is included.
(PDF)

S3 Table. Other primer sets used in this study. Includes primer pairs used for RT-PCR, RT-qPCR, and arbitrary PCR.
(PDF)

S1 Dataset. Analysis of RqlH and RqlI distribution and synteny. Contains a list of RqlH and RqlI orthologous proteins and the genomic locations of their respective genes. Quantification of the various genera that contain RqlH and RqlI orthologs, as well as a list of all of the genera included in this analysis, are also presented.
(XLSX)

S1 Fig. Activation of the SOS response in WT F11 and F11Δ*rqlI* under varying oxygen conditions. F11 (WT) or F11Δ*rqlI* (Δ) were grown in modified M9 media aerobically for 4 h or for 24 h under microaerobic or anaerobic conditions. Upper western blot shows levels of GFP expressed from strains carrying the SOS activation reporter construct P_{sulA}-GFP ("GFP"; pJLJ3) or the empty vector control containing a promoterless GFP ("EV"; pJLJ1). Below is the same blot probed with anti-*E. coli* antibody, presented as a loading control.
(TIF)

S2 Fig. Alignment of the RqlH proteins of F11 and *Mycobacterium smegmatis*. F11 RqlH (EcF11_3932) and M. smegmatis RqlH (MSMEG_5935) were aligned by EMBOSS Needle using the default settings. The protein alignment demonstrates that these two proteins have 41.2% identity and 55.7% similarity. The shading indicates various domains as delineated by Ordonez et al [1] and as indicated in the legend at right.
(TIF)

S3 Fig. Overexpression of RqlH reduces bacterial growth, especially in the absence of RqlI. Wild type (WT), Δ*rqlH*, Δ*rqlI*, and Δ*rqlHI* cells carried either a control plasmid (pBAD33) or an arabinose-inducible rqlH expression plasmid (pCWR21). The cells were grown overnight in M9 with chloramphenicol to maintain the plasmids, and then were subcultured 1:100 into fresh M9 with chloramphenicol. The bacteria were then grown either aerobically for 4 h, or microaerobically for 6 h, at which time the cultures were titered to determine bacterial concentration. Arabinose (0.05%) was present in all cultures, added either at the time of subculturing (microaerobic), or after 2 h of growth (aerobic). Graphs show mean values ± SEM from three experiments performed in triplicate. *, $P \leq 0.05$; ****, $P \leq 0.0001$ as determined by Student's *t* test.
(TIF)

S4 Fig. Expression of WT and mutant RqlH proteins. MG1655*attTn7*::P_{sulA}-GFP strains carrying the indicated RqlH expression plasmids or the empty vector pBAD33 were grown aerobically for 4 h in modified M9 media and then processed for western blot analysis. Upper blot was probed with anti-HA antibody to detect the epitope-tagged RqlH variants. *, non-specific background band present in all lanes. At the bottom is the same blot probed with anti-*E. coli* antibody, presented as a loading control.
(TIF)

S5 Fig. The RqlH PRTase domain is responsible for loss of cell viability in the absence of RqlI. MG1655 strains carrying the empty vector pRR48 or the RqlI expression plasmid pCWR16 were transformed with pBAD33-based constructs encoding full-length RqlH or various RqlH variants, as indicated. See Fig 5 for a schematic of the WT and mutant RqlH proteins that are expressed. Strains were diluted from overnight cultures and grown aerobically in modified M9 media. After 6 h, the (A) density (OD_{600}) and (B) number of viable bacteria (CFU/mL) were determined and normalized to the control strain (EV) carrying the empty vectors pRR48 and pBAD33. Graphs show mean values ± SEM from three experiments performed in triplicate. #, $P<0.05$ versus EV controls, as determined by Student's *t* test. (C) Following the 6 h growth period, each recombinant strain was diluted 1:100 into fresh modified M9 media and aerobic growth in 100-well honeycomb plates was assessed using a Bioscreen C instrument. The time that it took each culture to reach an OD_{600} of 0.4 is graphed versus the $\log_{10}$-transformed data from (B). These assays were carried out in media with added IPTG and arabinose to induce expression of the recombinant RqlI and RqlH proteins.
(TIF)

S6 Fig. The RqlH PRTase domain is necessary to induce filamentation and the SOS response in the absence of RqlI. MG1655 *attTn7*::P_{sulA}-GFP carrying the empty vector pRR48 or the RqlI expression plasmid pCWR16 along with the indicated WT and mutant RqlH expression constructs were grown aerobically for 6 h in modified M9 media. IPTG and arabinose were included to induce expression of the recombinant RqlI and RqlH proteins, respectively. See Fig 5 for a schematic of the RqlH proteins that are expressed. (A) Bacterial cells were imaged and measured using ImageJ. Boxes indicate median and interquartile ranges with whiskers extending to the 1st and 99th percentiles. (B) Western blots indicate levels of GFP expression downstream of the *sulA* promoter in the recombinant strains ± RqlI, as indicated. As a loading control, blots were stripped and re-probed using anti-*E. coli* antibody.
(TIF)

S7 Fig. Putative active site residues within the PRTase domain of RqlH. (A) PurF, Hpt, and the PRTase domain of RqlH were aligned using T-Coffee [1]. Regions corresponding to the active sites of the PurF and Hpt PRTases are shown with residues known to interact with a nucleotide highlighted in red [2,3]. Residues within RqlH that are predicted to interact with a nucleotide are highlighted in green. (B) A portion of the crystal structure of PurF showing the nucleotide-binding pocket. Adenosine monophosphate (AMP) is highlighted in green, while residues that interact with AMP are red. (C) The predicted structure of the RqlH PRTase domain generated by Phyre2 [4] was overlaid with the crystal structure of PurF. The putative nucleotide-binding pocket of RqlH is shown, in the same orientation as PurF AMP-binding pocket depicted in (B). Residues within RqlH that are predicted to be important for PRTase activity are highlighted in green (C).
(TIF)

S8 Fig. Mutation of RqlH does not result in the development of anucleate bacteria. MG1655 *attTn7*::P_{sulA}-GFP carrying pRR48 and either pBAD33 (empty vector, EV) or pCWR23 (for expression of the RqlH K49A mutant protein) were grown aerobically for 4 h in the presence of arabinose for the induction of RqlH expression. Bacterial DNA was stained with Hoechst dye, and samples were imaged by bright field and fluorescence microscopy. Representative images are shown. In the merged images, the DNA signal is false-colored yellow.
(TIF)

S9 Fig. Expression of recombinant full-length and truncated RqlI variants. F11Δ*rqlI* was transformed with the empty vector pRR48 or with plasmids encoding FLAG-tagged full-length RqlI (pCWR16) or different RqlI truncation mutants, as indicated. See Fig 6 for a schematic of the RqlI variants examined. Strains were grown microaerobically for 24 h in modified M9 media. Ampicillin and IPTG were included for plasmid maintenance and for induction of RqlI expression, respectively. Upper blot shows levels of RqlI and its variants, detected using anti-FLAG antibody. *, non-specific background band. Below, as a loading control, is the same blot probed using anti-*E. coli* antibody.
(TIF)

S10 Fig. The putative HTH domain of RqlI is dispensable, while the N-terminus is required for inhibiting RqlH toxicity. (A) WT F11 or F11Δ*rqlI* carrying empty vector (EV, pRR48) or constructs for expression of full-length and RqlI truncation mutants were grown microaerobically in modified M9 media for 24 h before titering. See Fig 6 for a schematic of the RqlI variants that were tested. Graph shows mean titers ± SEM from three experiments performed in triplicate. *, $P<0.05$ versus F11Δ*rqlI* control, as determined by Student's *t* test. (B) Cultures of MG1655 expressing different RqlI variants with or without RqlH, as indicated, were grown aerobically for 6 h and then titered. Bars indicate mean titers ± SEM from three experiments

performed in triplicate. *, P<0.05 versus MG1655 with empty vector (EV) control plasmids, as determined by Student's *t* test. In these experiments, arabinose and IPTG were added to the media to induce expression of the RqlH and the RqlI variants, respectively.
(TIF)

S11 Fig. RqlI is required for gut colonization by the ExPEC isolates CFT073 and 536. Adult female Balb/c mice were inoculated via oral gavage with $1x10^9$ bacteria comprised of a 1:1 mix of (A) WT CFT073 plus CFT073Δ*rqlI* or (B) WT 536 plus 536 Δ*rqlI*. Graphs show competitive indices (CI), as calculated by dilution plating of fecal homogenates on selective media at the indicated time points. Bars represent median values; n = 8 to 10 mice.
(TIF)

S12 Fig. Persister cell development by WT F11 and F11Δ*rqlI*. Graph shows the percentage of surviving bacteria present 5 h after addition of ampicillin (100 mg/mL) or ciprofloxacin (10 mg/mL) to cultures in log-phase growth. Percent survival was calculated by titering cultures before and after antibiotic treatments. Data represent mean results ± SD from three independent experiments. *, P<0.05 as determined by Student's *t* test.
(TIF)

Acknowledgments

We acknowledge the assistance of Brittany Fleming and Sydney Wilson with some of the microbiological assays, as well as the help of Justin Panich with cloning. We thank the members of the University of Utah Office of Comparative Medicine for assistance with animal husbandry, as well as the University of Utah DNA Sequencing core. We are indebted to the other members of the Mulvey lab, Matthew Bochman, and Jessica Brown for enlightening scientific discussions.

Author Contributions

Conceived and designed the experiments: CWR MAM. Performed the experiments: CWR. Analyzed the data: CWR MAM. Contributed reagents/materials/analysis tools: CWR MAM. Wrote the paper: CWR MAM.

References

1. Barber AE, Norton JP, Spivak AM, Mulvey MA (2013) Urinary tract infections: current and emerging management strategies. Clin Infect Dis 57: 719–724. doi: 10.1093/cid/cit284 PMID: 23645845
2. Wiles TJ, Kulesus RR, Mulvey MA (2008) Origins and virulence mechanisms of uropathogenic Escherichia coli. Exp Mol Pathol 85: 11–19. doi: 10.1016/j.yexmp.2008.03.007 PMID: 18482721
3. Foxman B, Barlow R, D'Arcy H, Gillespie B, Sobel JD (2000) Urinary tract infection: self-reported incidence and associated costs. Ann Epidemiol 10: 509–515. PMID: 11118930
4. Foxman B (2010) The epidemiology of urinary tract infection. Nat Rev Urol 7: 653–660. doi: 10.1038/nrurol.2010.190 PMID: 21139641
5. Laupland KB (2013) Incidence of bloodstream infection: a review of population-based studies. Clin Microbiol Infect 19: 492–500. doi: 10.1111/1469-0691.12144 PMID: 23398633
6. Mylotte JM, Tayara A, Goodnough S (2002) Epidemiology of bloodstream infection in nursing home residents: evaluation in a large cohort from multiple homes. Clin Infect Dis 35: 1484–1490. PMID: 12471567
7. Siegman-Igra Y, Fourer B, Orni-Wasserlauf R, Golan Y, Noy A, et al. (2002) Reappraisal of community-acquired bacteremia: a proposal of a new classification for the spectrum of acquisition of bacteremia. Clin Infect Dis 34: 1431–1439. PMID: 12015688

8. Hilali F, Ruimy R, Saulnier P, Barnabe C, Lebouguenec C, et al. (2000) Prevalence of virulence genes and clonality in Escherichia coli strains that cause bacteremia in cancer patients. Infect Immun 68: 3983–3989. PMID: 10858212

9. Simonsen KA, Anderson-Berry AL, Delair SF, Davies HD (2014) Early-onset neonatal sepsis. Clin Microbiol Rev 27: 21–47. doi: 10.1128/CMR.00031-13 PMID: 24396135

10. Nicolas-Chanoine MH, Bertrand X, Madec JY (2014) Escherichia coli ST131, an intriguing clonal group. Clin Microbiol Rev 27: 543–574. doi: 10.1128/CMR.00125-13 PMID: 24982321

11. Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, et al. (2015) Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. Genome Res 25: 119–128. doi: 10.1101/gr.180190.114 PMID: 25373147

12. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, et al. (2008) The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol 190: 6881–6893. doi: 10.1128/JB.00619-08 PMID: 18676672

13. van Opijnen T, Lazinski DW, Camilli A (2015) Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms. Curr Protoc Microbiol 36: 1E 3 1–1E 3 24.

14. Wiles TJ, Norton JP, Russell CW, Dalley BK, Fischer KF, et al. (2013) Combining quantitative genetic footprinting and trait enrichment analysis to identify fitness determinants of a bacterial pathogen. PLoS Genet 9: e1003716. doi: 10.1371/journal.pgen.1003716 PMID: 23990803

15. Schnoes AM, Brown SD, Dodevski I, Babbitt PC (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput Biol 5: e1000605. doi: 10.1371/journal.pcbi.1000605 PMID: 20011109

16. van Opijnen T, Camilli A (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome Res 22: 2541–2551. doi: 10.1101/gr.137430.112 PMID: 22826510

17. Deutschbauer A, Price MN, Wetmore KM, Tarjan DR, Xu Z, et al. (2014) Towards an informative mutant phenotype for every bacterial gene. J Bacteriol 196: 3643–3655. doi: 10.1128/JB.01836-14 PMID: 25112473

18. Kisker C, Schindelin H, Rees DC (1997) Molybdenum-cofactor-containing enzymes: structure and mechanism. Annu Rev Biochem 66: 233–267. PMID: 9242907

19. Karudapuram S, Zhao X, Barcak GJ (1995) DNA sequence and characterization of Haemophilus influenzae dprA+, a gene required for chromosomal but not plasmid DNA transformation. J Bacteriol 177: 3235–3240. PMID: 7768823

20. Janion C (2008) Inducible SOS response system of DNA repair and mutagenesis in Escherichia coli. Int J Biol Sci 4: 338–344. PMID: 18825275

21. Ordonez H, Unciuleac M, Shuman S (2012) Mycobacterium smegmatis RqlH defines a novel clade of bacterial RecQ-like DNA helicases with ATP-dependent 3'-5' translocase and duplex unwinding activities. Nucleic Acids Res 40: 4604–4614. doi: 10.1093/nar/gks046 PMID: 22287622

22. Luria SE, Delbruck M (1943) Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics 28: 491–511. PMID: 17247100

23. Jones HM, Gunsalus RP (1985) Transcription of the Escherichia coli fumarate reductase genes (frdABCD) and their coordinate regulation by oxygen, nitrate, and fumarate. J Bacteriol 164: 1100–1109. PMID: 2999070

24. Yampolsky LY, Stoltzfus A (2005) The exchangeability of amino acids in proteins. Genetics 170: 1459–1472. PMID: 15944362

25. Wu Y, Brosh RM Jr. (2010) Helicase-inactivating mutations as a basis for dominant negative phenotypes. Cell Cycle 9: 4080–4090. doi: 10.4161/cc.9.20.13667 PMID: 20980836

26. Albenberg L, Esipova TV, Judge CP, Bittinger K, Chen J, et al. (2014) Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. Gastroenterology 147: 1055–1063 e1058. doi: 10.1053/j.gastro.2014.07.020 PMID: 25046162

27. Hagan EC, Lloyd AL, Rasko DA, Faerber GJ, Mobley HL (2010) Escherichia coli global gene expression in urine from women with urinary tract infection. PLoS Pathog 6: e1001187. doi: 10.1371/journal.ppat.1001187 PMID: 21085611

28. Croteau DL, Popuri V, Opresko PL, Bohr VA (2014) Human RecQ helicases in DNA repair, recombination, and replication. Annu Rev Biochem 83: 519–552. doi: 10.1146/annurev-biochem-060713-035428 PMID: 24606147

29. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, et al. (2005) Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A 102: 11070–11075. PMID: 16033867

30. Krych L, Hansen CH, Hansen AK, van den Berg FW, Nielsen DS (2013) Quantitatively different, yet qualitatively alike: a meta-analysis of the mouse core gut microbiome with a view towards the human gut microbiome. PLoS One 8: e62578. doi: 10.1371/journal.pone.0062578 PMID: 23658749

31. Holden JA, Harriman PD, Wall JD (1976) Escherichia coli mutants deficient in guanine-xanthine phosphoribosyltransferase. J Bacteriol 126: 1141–1148. PMID: 820683

32. Mortier-Barriere I, Velten M, Dupaigne P, Mirouze N, Pietrement O, et al. (2007) A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell 130: 824–836. PMID: 17803906

33. Smeets LC, Becker SC, Barcak GJ, Vandenbroucke-Grauls CM, Bitter W, et al. (2006) Functional characterization of the competence protein DprA/Smf in Escherichia coli. FEMS Microbiol Lett 263: 223–228. PMID: 16978360

34. Londono-Vallejo JA, Dubnau D (1993) comF, a Bacillus subtilis late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. Mol Microbiol 9: 119–131. PMID: 8412657

35. Tomb JF, el-Hajj H, Smith HO (1991) Nucleotide sequence of a cluster of genes involved in the transformation of Haemophilus influenzae Rd. Gene 104: 1–10. PMID: 1916268

36. Etchuuya R, Ito M, Kitano S, Shigi F, Sobue R, et al. (2011) Cell-to-cell transformation in Escherichia coli: a novel type of natural transformation involving cell-derived DNA and a putative promoting pheromone. PLoS One 6: e16355. doi: 10.1371/journal.pone.0016355 PMID: 21283723

37. Sun D, Zhang Y, Mei Y, Jiang H, Xie Z, et al. (2006) Escherichia coli is naturally transformable in a novel transformation system. FEMS Microbiol Lett 265: 249–255. PMID: 17069625

38. Sinha S, Redfield RJ (2012) Natural DNA uptake by Escherichia coli. PLoS One 7: e35620. doi: 10.1371/journal.pone.0035620 PMID: 22532864

39. Unterholzner SJ, Poppenberger B, Rozhon W (2013) Toxin-antitoxin systems: Biology, identification, and application. Mob Genet Elements 3: e26219. PMID: 24251069

40. Norton JP, Mulvey MA (2012) Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic Escherichia coli. PLoS Pathog 8: e1002954. doi: 10.1371/journal.ppat.1002954 PMID: 23055930

41. Sberro H, Leavitt A, Kiro R, Koh E, Peleg Y, et al. (2013) Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. Mol Cell 50: 136–148. doi: 10.1016/j.molcel.2013.02.002 PMID: 23478446

42. Murphy KC, Campellone KG (2003) Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic E. coli. BMC Mol Biol 4: 11. PMID: 14672541

43. Li XT, Thomason LC, Sawitzke JA, Costantino N, Court DL (2013) Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. Nucleic Acids Res 41: e204. doi: 10.1093/nar/gkt1075 PMID: 24203710

44. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640–6645. PMID: 10829079

45. Studdert CA, Parkinson JS (2005) Insights into the organization and dynamics of bacterial chemoreceptor clusters through in vivo crosslinking studies. Proc Natl Acad Sci U S A 102: 15623–15628. PMID: 16230637

46. Guzman LM, Belin D, Carson MJ, Beckwith J (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol 177: 4121–4130. PMID: 7608087

47. Bryksin A, Matsumura I (2013) Overlap extension PCR cloning. Methods Mol Biol 1073: 31–42. doi: 10.1007/978-1-62703-625-2_4 PMID: 23996437

48. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9: 671–675. PMID: 22930834

49. Zhang X, Bremer H (1995) Control of the Escherichia coli rrnB P1 promoter strength by ppGpp. J Biol Chem 270: 11181–11189. PMID: 7538113

50. Hall BM, Ma CX, Liang P, Singh KK (2009) Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics 25: 1564–1565. doi: 10.1093/bioinformatics/btp253 PMID: 19369502

51. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25: 1422–1423. doi: 10.1093/bioinformatics/btp163 PMID: 19304878

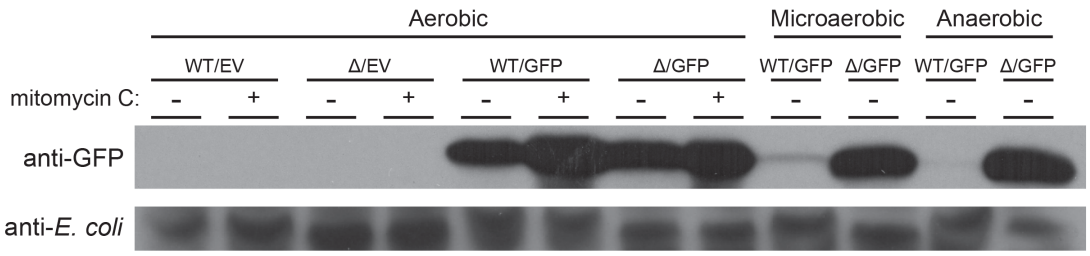


Figure 3.S1. Activation of the SOS response in WT F11 and F11Δ*rqlI* under varying oxygen conditions. F11 (WT) or F11Δ*rqlI* (Δ) were grown in modified M9 media aerobically for 4 h or for 24 h under microaerobic or anaerobic conditions. Upper western blot shows levels of GFP expressed from strains carrying the SOS activation reporter construct P_{sulA}-GFP ("GFP"; pJLJ3) or the empty vector control containing a promoterless GFP ("EV"; pJLJ1). Below is the same blot probed with anti-*E. coli* antibody, presented as a loading control.

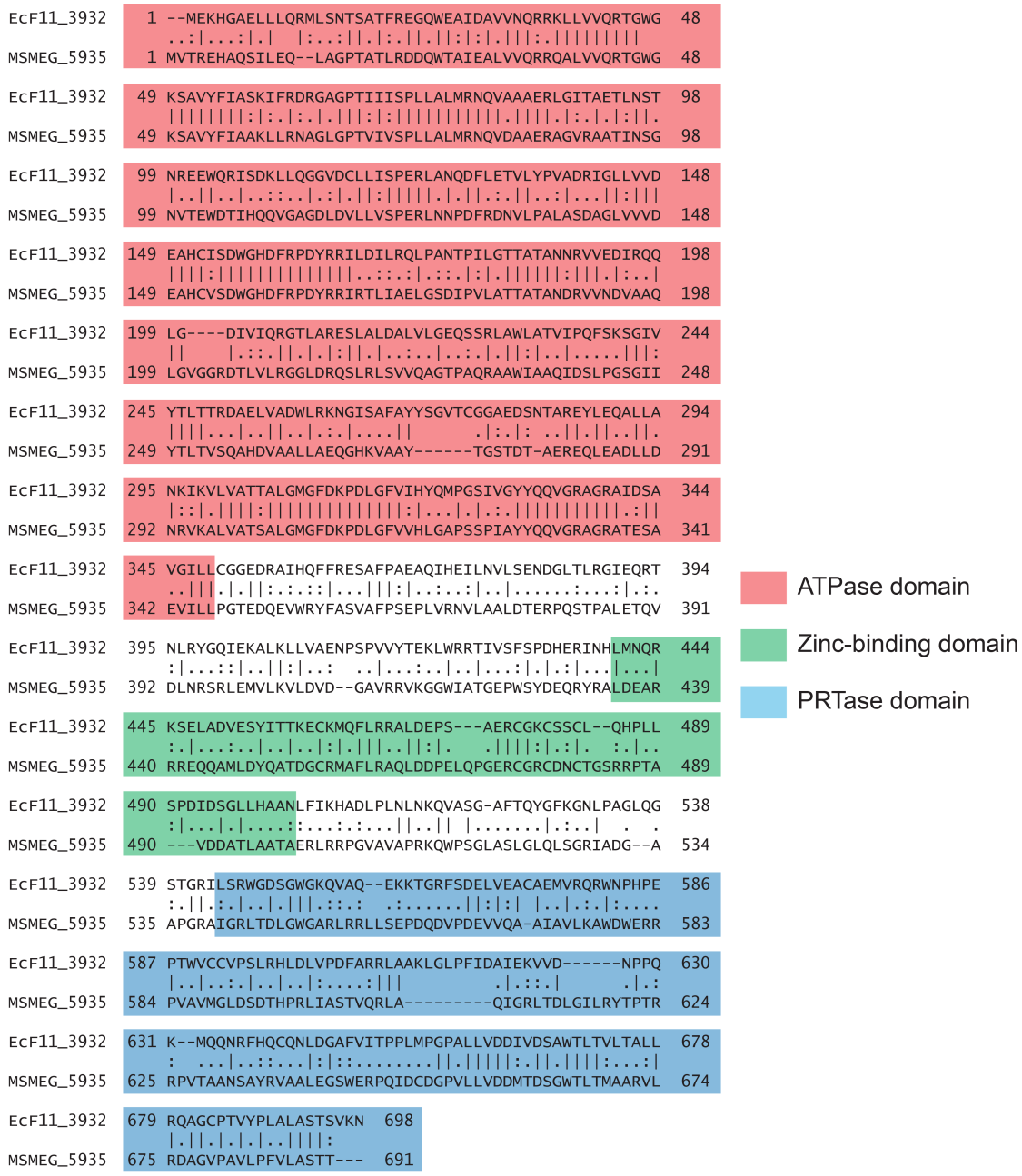


Figure 3.S2. Alignment of the RqlH proteins of F11 and *Mycobacterium smegmatis*. F11 RqlH (EcF11_3932) and M. smegmatis RqlH (MSMEG_5935) were aligned by EMBOSS Needle using the default settings. The protein alignment demonstrates that these two proteins have 41.2% identity and 55.7% similarity. The shading indicates various domains as delineated by Ordonez et al and as indicated in the legend at right.

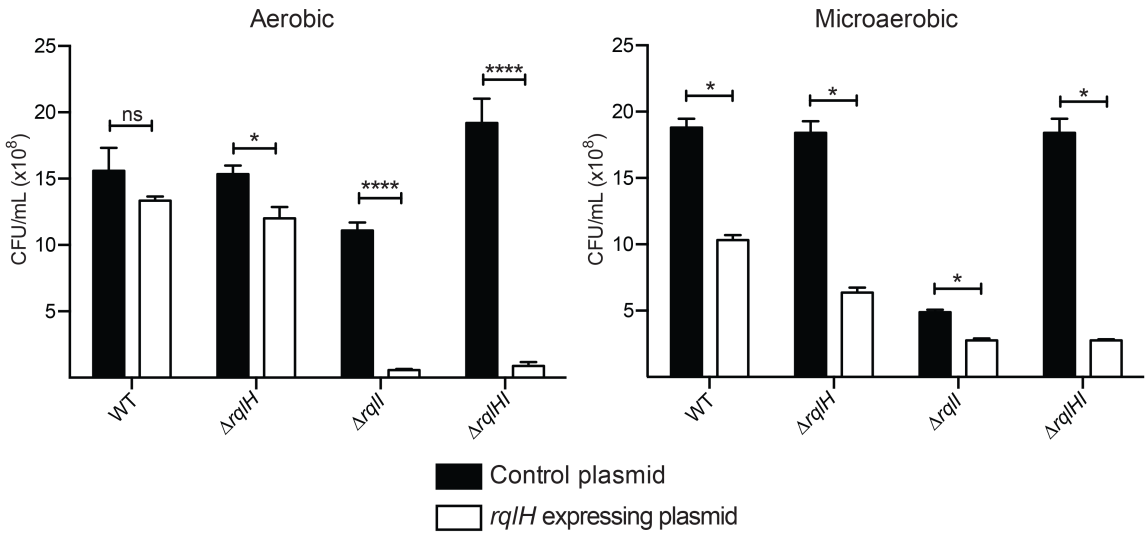


Figure 3.S3. Overexpression of RqlH reduces bacterial growth, especially in the absence of RqlI. Wild type (WT), Δ*rqlH*, Δ*rqlI*, and Δ*rqlHI* cells carried either a control plasmid (pBAD33) or an arabinose-inducible rqlH expression plasmid (pCWR21). The cells were grown overnight in M9 with chloramphenicol to maintain the plasmids, and then were subcultured 1:100 into fresh M9 with chloramphenicol. The bacteria were then grown either aerobically for 4 h, or microaerobically for 6 h, at which time the cultures were titered to determine bacterial concentration. Arabinose (0.05%) was present in all cultures, added either at the time of subculturing (microaerobic), or after 2 h of growth (aerobic). Graphs show mean values ± SEM from three experiments performed in triplicate. *, $P \leq 0.05$; ****, $P \leq 0.0001$ as determined by Student's *t* test.

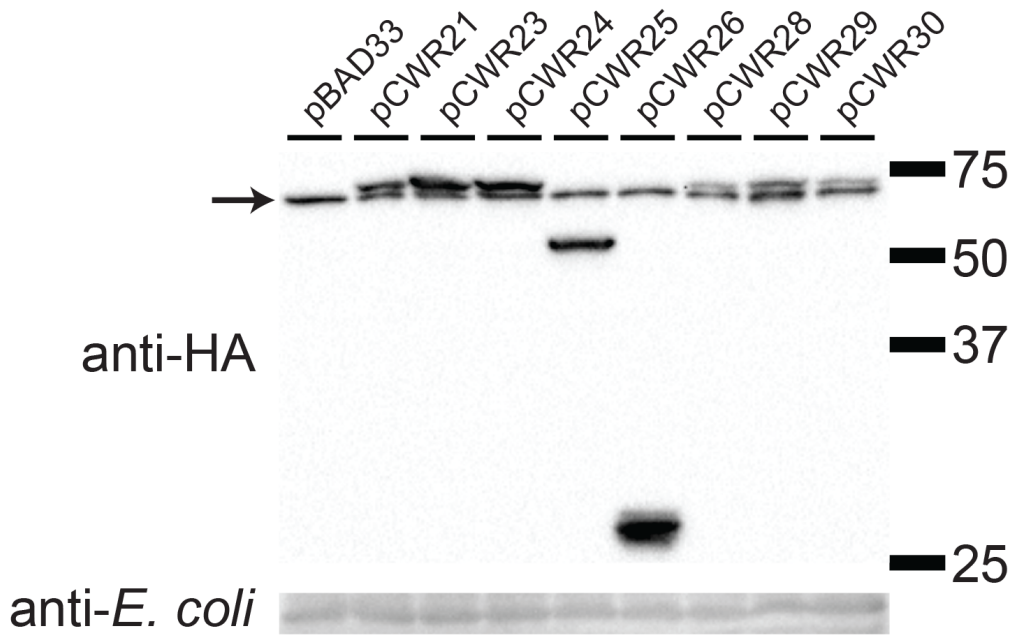


Figure 3.S4. Expression of WT and mutant RqlH proteins.
MG1655*attTn7*::P_{sulA}-GFP strains carrying the indicated RqlH expression plasmids or the empty vector pBAD33 were grown aerobically for 4 h in modified M9 media and then processed for western blot analysis. Upper blot was probed with anti-HA antibody to detect the epitope-tagged RqlH variants. *, non-specific background band present in all lanes. At the bottom is the same blot probed with anti-*E. coli* antibody, presented as a loading control.

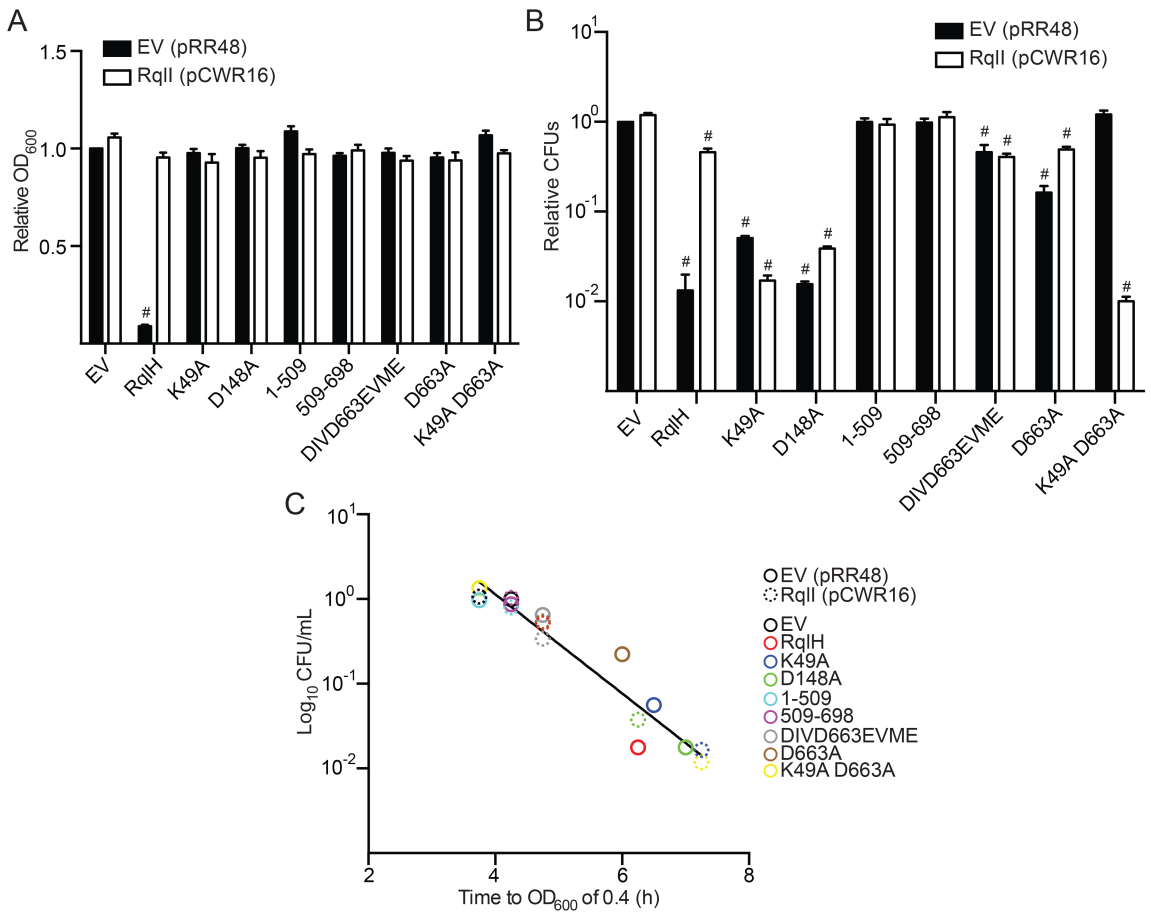


Figure 3.S5. The RqlH PRTase domain is responsible for loss of cell viability in the absence of RqlI. MG1655 strains carrying the empty vector pRR48 or the RqlI expression plasmid pCWR16 were transformed with pBAD33-based constructs encoding full-length RqlH or various RqlH variants, as indicated. See Fig 5 for a schematic of the WT and mutant RqlH proteins that are expressed. Strains were diluted from overnight cultures and grown aerobically in modified M9 media. After 6 h, the (A) density (OD_{600}) and (B) number of viable bacteria (CFU/mL) were determined and normalized to the control strain (EV) carrying the empty vectors pRR48 and pBAD33. Graphs show mean values ± SEM from three experiments performed in triplicate. #, $P<0.05$ versus EV controls, as determined by Student's *t* test. (C) Following the 6 h growth period, each recombinant strain was diluted 1:100 into fresh modified M9 media and aerobic growth in 100-well honeycomb plates was assessed using a Bioscreen C instrument. The time that it took each culture to reach an OD_{600} of 0.4 is graphed versus the $\log_{10}$-transformed data from (B). These assays were carried out in media with added IPTG and arabinose to induce expression of the recombinant RqlI and RqlH proteins.

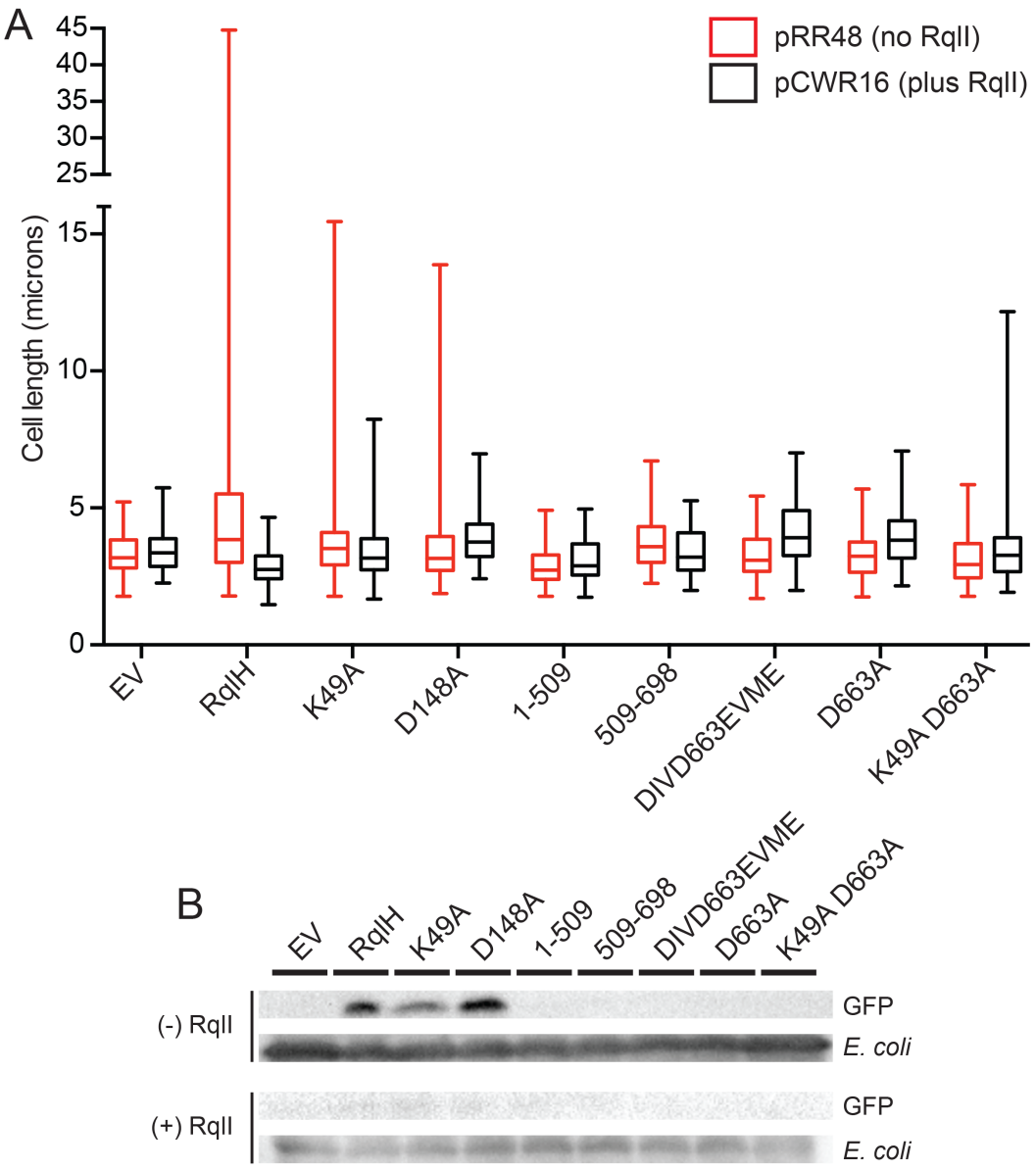


Figure 3.S6. The RqlH PRTase domain is necessary to induce filamentation and the SOS response in the absence of RqlI. MG1655 *attTn7*::P_{sulA}-GFP carrying the empty vector pRR48 or the RqlI expression plasmid pCWR16 along with the indicated WT and mutant RqlH expression constructs were grown aerobically for 6 h in modified M9 media. IPTG and arabinose were included to induce expression of the recombinant RqlI and RqlH proteins, respectively. See Fig 5 for a schematic of the RqlH proteins that are expressed. (A) Bacterial cells were imaged and measured using ImageJ. Boxes indicate median and interquartile ranges with whiskers extending to the 1st and 99th percentiles. (B) Western blots indicate levels of GFP expression downstream of the *sulA* promoter in the recombinant strains ± RqlI, as indicated. As a loading control, blots were stripped and re-probed using anti-*E. coli* antibody.

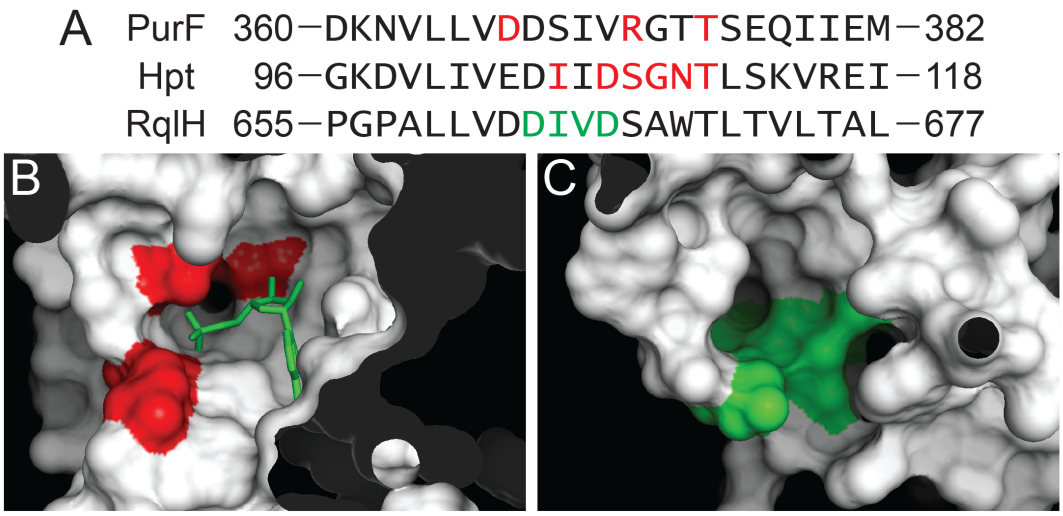


Figure 3.S7. Putative active site residues within the PRTase domain of RqlH. (A) PurF, Hpt, and the PRTase domain of RqlH were aligned using T-Coffee. Regions corresponding to the active sites of the PurF and Hpt PRTases are shown with residues known to interact with a nucleotide highlighted in red. Residues within RqlH that are predicted to interact with a nucleotide are highlighted in green. (B) A portion of the crystal structure of PurF showing the nucleotide-binding pocket. Adenosine monophosphate (AMP) is highlighted in green, while residues that interact with AMP are red. (C) The predicted structure of the RqlH PRTase domain generated by Phyre2 was overlaid with the crystal structure of PurF. The putative nucleotide-binding pocket of RqlH is shown, in the same orientation as PurF AMP-binding pocket depicted in (B). Residues within RqlH that are predicted to be important for PRTase activity are highlighted in green (C).

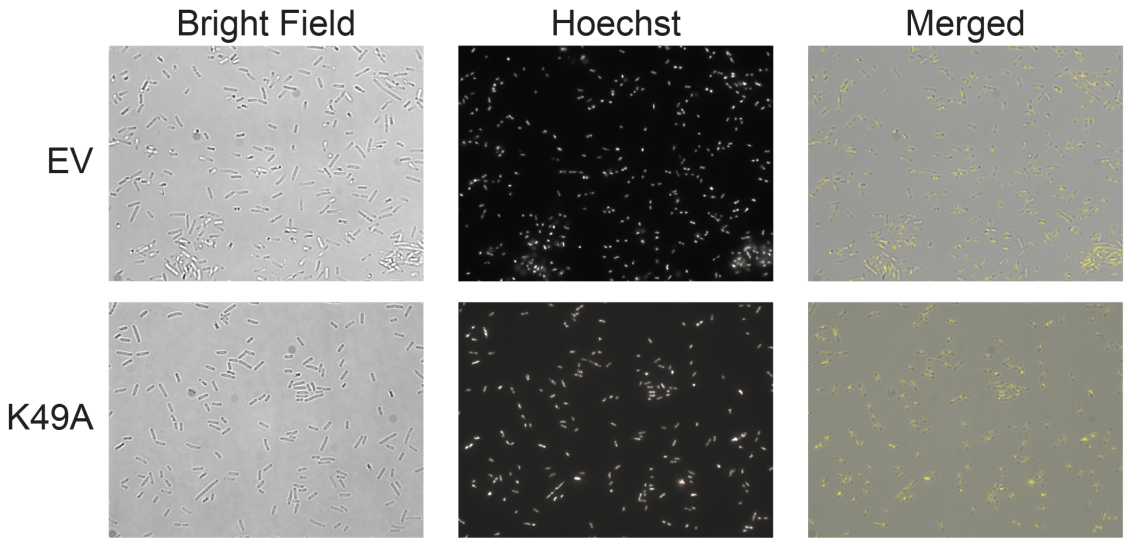


Figure 3.S8. Mutation of RqlH does not result in the development of anucleate bacteria. MG1655 *attTn7::*P_{sulA}-GFP carrying pRR48 and either pBAD33 (empty vector, EV) or pCWR23 (for expression of the RqlH K49A mutant protein) were grown aerobically for 4 h in the presence of arabinose for the induction of RqlH expression. Bacterial DNA was stained with Hoechst dye, and samples were imaged by bright field and fluorescence microscopy. Representative images are shown. In the merged images, the DNA signal is false-colored yellow.

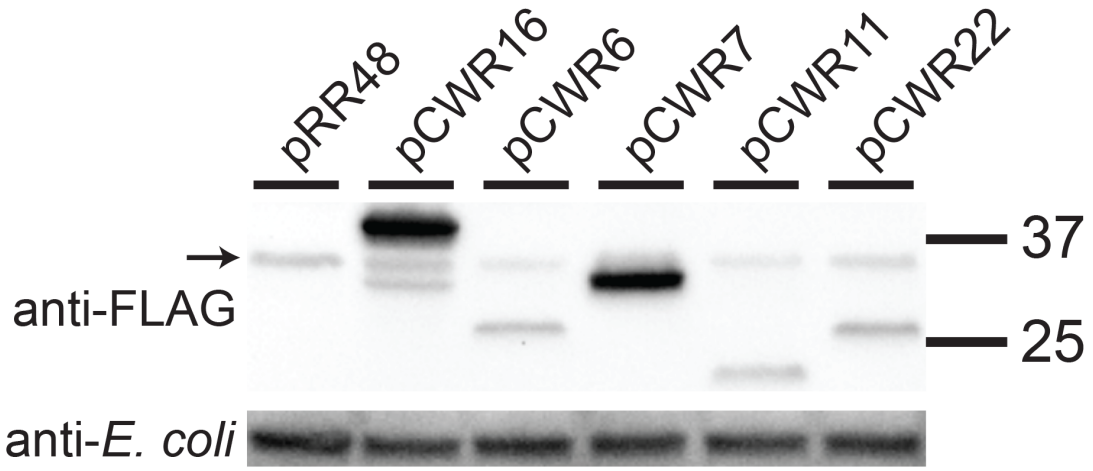


Figure 3.S9. Expression of recombinant full-length and truncated RqlI variants. F11Δ*rqlI* was transformed with the empty vector pRR48 or with plasmids encoding FLAG-tagged full-length RqlI (pCWR16) or different RqlI truncation mutants, as indicated. See Fig 6 for a schematic of the RqlI variants examined. Strains were grown microaerobically for 24 h in modified M9 media. Ampicillin and IPTG were included for plasmid maintenance and for induction of RqlI expression, respectively. Upper blot shows levels of RqlI and its variants, detected using anti-FLAG antibody. *, non-specific background band. Below, as a loading control, is the same blot probed using anti-*E. coli* antibody.

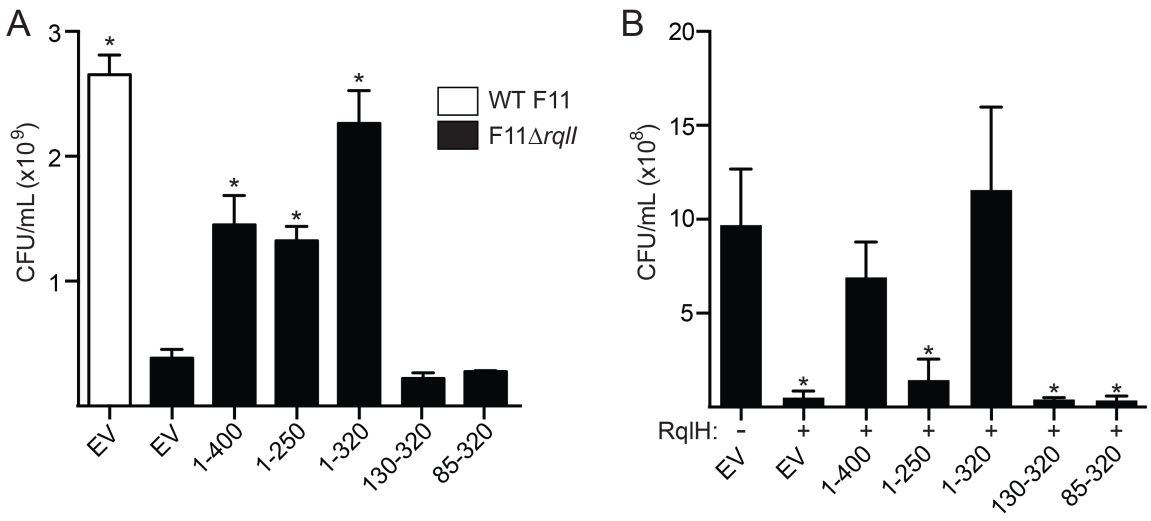


Figure 3.S10. The putative HTH domain of RqlI is dispensable, while the N-terminus is required for inhibiting RqlH toxicity. (A) WT F11 or F11Δ*rqlI* carrying empty vector (EV, pRR48) or constructs for expression of full-length and RqlI truncation mutants were grown microaerobically in modified M9 media for 24 h before titering. See Fig 6 for a schematic of the RqlI variants that were tested. Graph shows mean titers ± SEM from three experiments performed in triplicate. *, P<0.05 versus F11Δ*rqlI* control, as determined by Student's *t* test. (B) Cultures of MG1655 expressing different RqlI variants with or without RqlH, as indicated, were grown aerobically for 6 h and then titered. Bars indicate mean titers ± SEM from three experiments performed in triplicate. *, P<0.05 versus MG1655 with empty vector (EV) control plasmids, as determined by Student's *t* test. In these experiments, arabinose and IPTG were added to the media to induce expression of the RqlH and the RqlI variants, respectively.

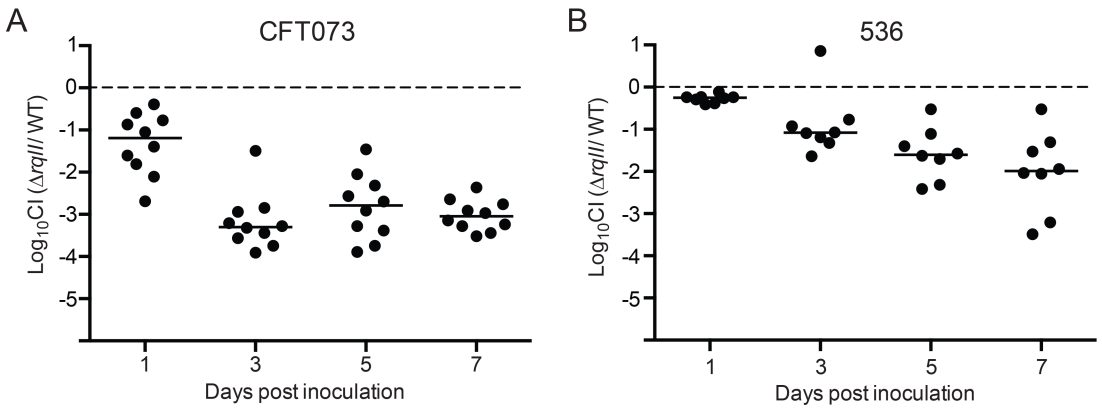


Figure 3.S11. RqlI is required for gut colonization by the ExPEC isolates CFT073 and 536. Adult female Balb/c mice were inoculated via oral gavage with 1x10^9 bacteria comprised of a 1:1 mix of (A) WT CFT073 plus CFT073Δ*rqlI* or (B) WT 536 plus 536 Δ*rqlI*. Graphs show competitive indices (CI), as calculated by dilution plating of fecal homogenates on selective media at the indicated time points. Bars represent median values; n=8 to 10 mice.

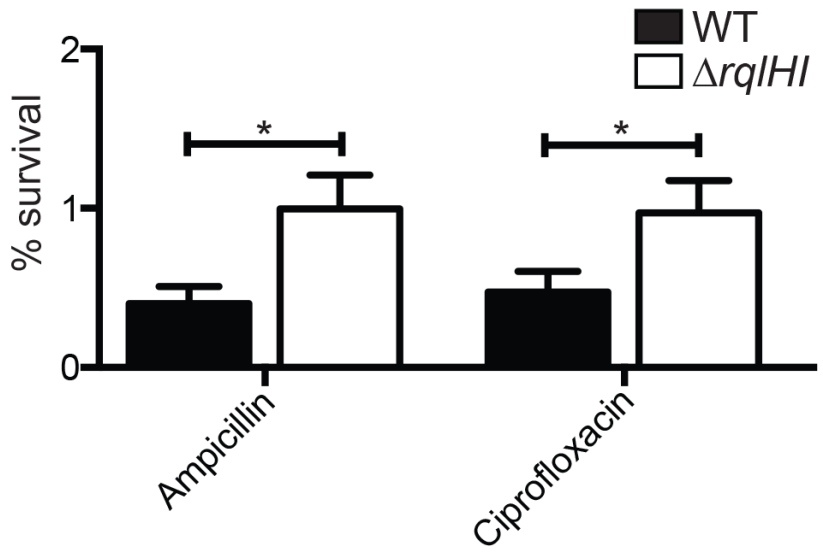


Figure 3.S12. Persister cell development by WT F11 and F11Δ*rqlI*. Graph shows the percentage of surviving bacteria present 5 h after addition of ampicillin (100 mg/mL) or ciprofloxacin (10 mg/mL) to cultures in log-phase growth. Percent survival was calculated by titering cultures before and after antibiotic treatments. Data represent mean results ± SD from three independent experiments. *, $P<0.05$ as determined by Student's *t* test.

CHAPTER 4

PROPER REGULATION OF THE GLYCEROL DEGRADATION PATHWAY IN EXTRAINTESTINAL PATHOGENIC ESCHERICHIA COLI IS REQUIRED FOR MUCUS METABOLISM AND GUT COLONIZATION

Colin W. Russell, Alexander S. Chang, Matthew A. Mulvey

Abstract

Extraintestinal pathogenic *Escherichia coli* (ExPEC) benignly colonizes the mammalian gut, but can disperse to extraintestinal sites to cause disease. As inhabitance of the intestine is often a prerequisite for ExPEC-mediated pathogenesis, we set out to understand how ExPEC colonizes this niche. A Tn-seq screen was performed to search for genes within the ExPEC isolate F11 that are important for growth in intestinal mucus, which is thought to be a source of nutrients for commensal *E. coli* in the gut. Several genes that contribute to fitness in mucus broth were discovered, with the beta-oxidation pathway being especially important, indicating that metabolism of mucus includes utilization of fatty acids. Some genes that were important for mucus metabolism, but that are not traditionally a part of beta-oxidation, were found to contribute in some fashion to degradation of fatty acids as they were important for growth on oleic acid, a long chain fatty acid. This included the *glpG* gene, whose mutation led to a disruption of the downstream gene *glpR*, the transcriptional repressor of glycerol degradation. The Δ*glpG* mutant exhibited a defect during mouse gut colonization, indicating that glycerol degradation must be carefully regulated *in vivo*. This work helps to elaborate the process by which ExPEC colonizes the mammalian gastrointestinal tract.

Introduction

Extraintestinal pathogenic *Escherichia coli* (ExPEC) is a group of *E. coli* strains that can colonize the mammalian gut without causing any overt pathology. However, ExPEC can spread from the gut to extraintestinal niches such as the urinary tract, the meninges, and the bloodstream, and induce disease [1]. Urinary tract infections (UTIs) are most often caused by ExPEC bacteria, are typically resolved with antibiotic treatment, and are usually not life-threatening, especially when infection is limited to the bladder [1]. However, UTIs are quite prevalent, with half of all women requiring

treatment during their lifetime [2], and many women suffering from recurring UTIs [3]. This prevalence contributes to a high economic burden, with over $2 billion spent treating bladder infections each year in the United States [4]. In addition to UTI, ExPEC can infect the bloodstream and cause severe disease, especially in patients with a weakened immune response. For example, in the context of early-onset neonatal sepsis, ExPEC are the second leading etiological agent and the primary cause of mortality [5]. The concern surrounding the spectrum and prevalence of ExPEC pathogenesis is compounded by the reality that antibiotic resistant strains are increasing in frequency and are spreading worldwide [6]. In order to combat ExPEC more efficiently, it is critical that we understand the ExPEC life cycle in greater depth.

There are many indications that the gut may act as a reservoir from which ExPEC can disseminate and cause disease. For example, during acute UTI, the same strain can frequently be isolated from both the patient's urine and feces [7-9], and in most of these instances, the uropathogen is the predominant fecal *E. coli* strain [7]. ExPEC can also translocate directly from the gut to the bloodstream, especially in cancer patients. Indeed, ExPEC is the most common Gram-negative bacteria isolated from the blood of cancer patients with bacteremia, who have a reduced gut barrier function due to immunosuppression and altered gut mucosa [10]. In addition to spreading from the gut to an extraintestinal site within a single individual, ExPEC can spread from one person to another, as in the case of neonatal sepsis, in which the newborn becomes infected by ExPEC originating from the mother [11]. Therefore, although ExPEC is not known to cause serious disease in the gastrointestinal tract, an understanding of the gut colonization process could be helpful for developing future therapeutic options to combat extraintestinal infections.

One element that may be key to gut colonization by ExPEC is the intestinal mucus layer, which functions to separate the host epithelium from the microbiota [12].

In addition to acting as a barrier, the mucus layer can be an important niche for the microbiota, through its supply of nutrients, and by acting as an attachment point for certain microbes [13]. The mucus layer appears to be one site of colonization for *E. coli* commensal strains, which can associate with the mucus layer in the streptomycin-treated mouse model of gut colonization [14, 15]. These *E. coli* commensal strains seem to prefer to utilize mucus as a source of nutrients, as they often grow *in vitro* in crude cecal mucus, but not in the cecal lumenal contents [15, 16]. There appears to be an association between the ability of a strain to grow in cecal mucus *in vitro*, and the ability of that strain to colonize the mouse gut [15-17], although this is not always a perfect correlation [18].

To better understand how ExPEC colonizes the gut, we set out to document the genes critical for growth in mucus *in vitro* by utilizing transposon sequencing (Tn-seq), a high-throughput method for determining the contribution of each bacterial gene to fitness under a given condition [19]. We found several genes that were important for growth in mucus, but not in the glucose control condition. Many of these mucus-specific genes were found to be important for proper metabolism of long chain fatty acids (LCFAs). One of these genes, which encodes GlpG, was found to be important for colonization of the mouse gut in the presence of the natural microbiota. The defect of the Δ*glpG* mutant strain is shown to be due to polar effects on *glpR*, the transcriptional repressor of the glycerol degradation pathway. These data demonstrate the importance of proper regulation of the glycerol degradation pathway during ExPEC colonization of the mouse gut, and point to a role for fatty acid degradation in mucus metabolism.

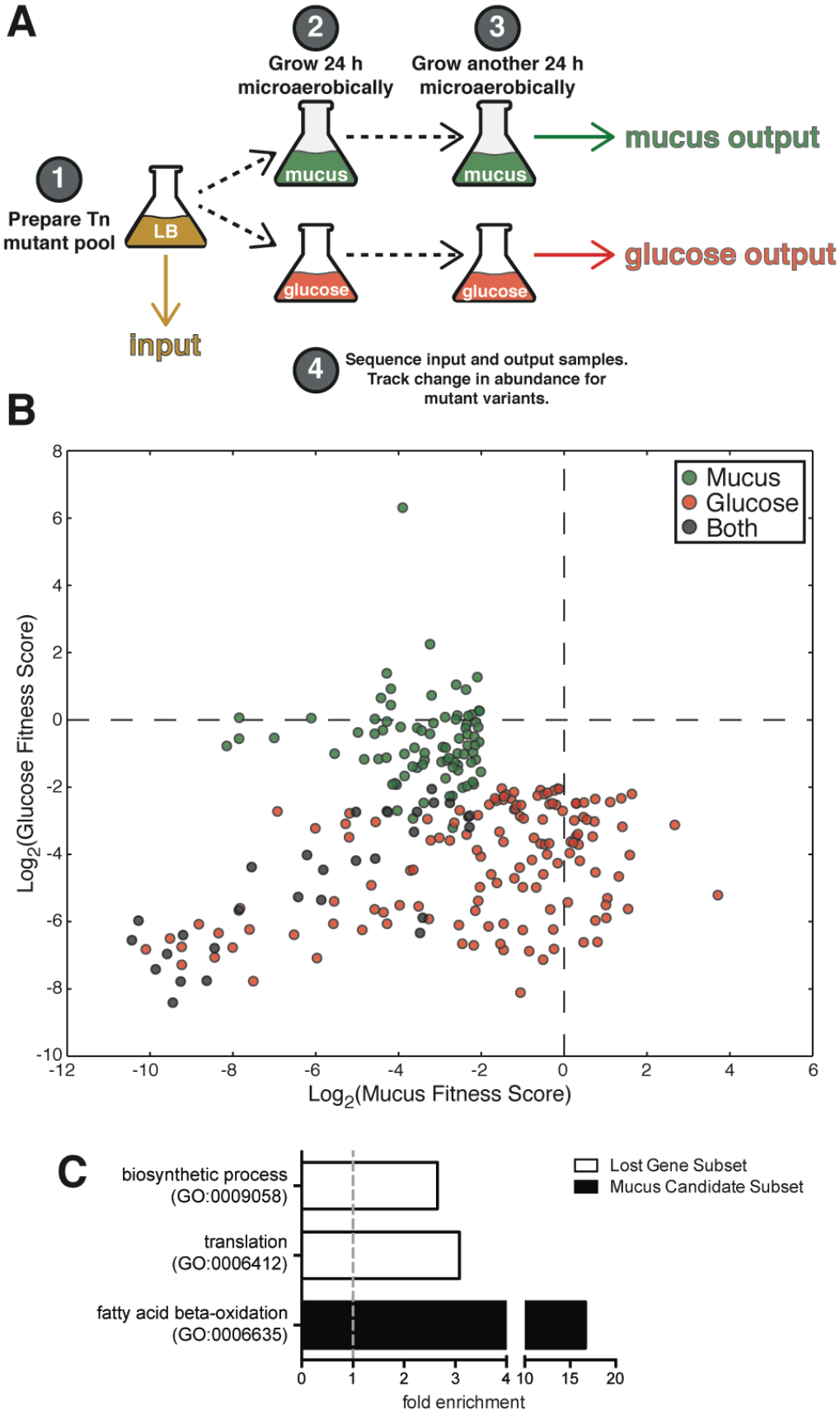
Results

A Tn-seq screen for ExPEC genes important for growth on intestinal mucus uncovers a role for beta-oxidation

As *E. coli* has often been observed to associate with gastrointestinal mucus [14, 15], and the mucus layer is known to be a source of nutrients for the microbiota [13], we reasoned that ExPEC genes required for metabolism of intestinal mucus may also be required for gut colonization. Therefore, a Tn-seq screen was devised in order to uncover genes important for mucus metabolism *in vitro* (Fig 4.1A). Three independently-derived mutant pools [20] were first expanded in LB, then subcultured into minimal media supplemented with either mucus or glucose. The bacteria were then grown for 24 h microaerobically to mimic the gut environment. A second round of 24 h of growth was performed in order to increase the stringency of the screen. After a total of 48 h of microaerobic growth in minimal media, the mutant populations were collected and the transposon insertion sites were sequenced in parallel with the input population. The abundance of individual mutants in the output and input pools were then compared, and a fitness score was derived for each mutant. After normalization and filtering (see Materials and Methods), each fitness score was calculated as the Log_2(mutant abundance in output/mutant abundance in input). The median value of the various insertion site fitness scores for a given gene was defined as the fitness score for that gene.

There were a total of 24,822 transposon insertion sites for which a fitness score was determined in both the mucus and glucose conditions. On average, there was one transposon insertion detected for every 210 nucleotides. This coverage allowed a fitness score to be calculated for 3,804 genes. In order to create a list of candidate genes that contribute to fitness in either the mucus media, the glucose media, or both, several cutoffs were used. Only genes with a fitness score of -2.0 or lower (i.e. a minimum of a 4-fold reduction of mutant abundance) and a P value less than 0.05 were included in the

Figure 4.1. Tn-seq uncovers a role for beta-oxidation in metabolizing mucus. (A) A schematic describing the Tn-seq experiment. An F11 transposon library was expanded aerobically in LB overnight, then subcultured 1:100 into a modified M9 medium supplemented with either mucus or glucose. The bacteria were then grown for 24 h microaerobically, at which point they were subcultured 1:100 into fresh media and grown for another 24 h microaerobically. This procedure was repeated for 3 independently-derived transposon mutant pools. Samples of bacterial cells were taken from the LB broth at the beginning of the experiment ("input"), and from the mucus and glucose media after 48 h of growth ("mucus output" and "glucose output"). These input and output samples were then prepared for sequencing of the transposon insertion sites, and the abundance of the various transposon mutants in all of the samples was quantified by bioinformatics. The abundance of each mutant in the output and input samples was used to derive a fitness score for that mutant. The mutant fitness scores for the three samples were pooled and used to calculate a fitness score for each gene in both the glucose and mucus conditions. (B) Gene fitness scores are graphed according to the gene's fitness score during growth in both the glucose and mucus broths. Only candidate genes are included in this graph, meaning that the gene fitness score was -2.0 or less (a 4-fold reduction in fitness) with $P \leq 0.05$. These thresholds were met in either the mucus condition (green dots), the glucose condition (red dots), or in both conditions (black dots). (C) Candidate genes were analyzed via GO analysis using PANTHER. Those genes that were not detected in the input sample ("Lost Gene Subset"), likely due to their essential nature, were also analyzed. Those GO categories that were significantly enriched in the Lost Gene Subset and the Mucus Candidate Subset when compared to the F11 genome are graphed.


A
2
Grow 24 h microaerobically
3
Grow another 24 h microaerobically
mucus output
1
Prepare Tn mutant pool
LB
mucus
mucus
glucose
glucose
glucose output
input
4
Sequence input and output samples. Track change in abundance for mutant variants.
B
Mucus
Glucose
Both
Log2(Glucose Fitness Score)
Log2(Mucus Fitness Score)
C
Lost Gene Subset
Mucus Candidate Subset
biosynthetic process (GO:0009058)
translation (GO:0006412)
fatty acid beta-oxidation (GO:0006635)
fold enrichment

candidate list (Fig 4.1B). These thresholds produced a list of 112 mucus-specific genes, 171 genes with fitness defects only in the glucose condition, and 28 genes that exhibited a loss of fitness in both conditions.

Several genes of note were found to be candidates in both the mucus and glucose media, likely because they are essential for microaerobiosis, a feature that both conditions had in common. For example, *rqlI* was found to be a candidate gene in both media, with a mucus score of -5.8 and a glucose score of -4.5. This is unsurprising as the Δ*rqlI* mutant exhibits a growth defect in low oxygen conditions due to the inherent toxicity of the helicase RqlH [21]. Another candidate gene important for low oxygen conditions was *arcB*, with a mucus score of -3.5 and a glucose score of -6.3. ArcB is a sensory histidine kinase that forms a two-component system with ArcA [22], a transcriptional regulator that is important for proper microaerobic metabolism [23]. In addition to *rqlI* and *arcB*, several proteins that are a part of NADH dehydrogenase I (NDH-1) were also candidate genes in both conditions, including *nuoF*, *nuoG*, and *nuoL*. Five other genes encoding NDH-1 components (*nuoC*, *nuoI*, *nuoJ*, *nuoM*, and *nuoN*) were candidates in the glucose condition, but had few or no transposon insertion mutants in the mucus broth, making it impossible to determine their importance in that medium. NDH-1 functions as part of the electron transport chain to transfer electrons from NADH to the quinone pool, and is especially efficient when fumarate or DMSO are being used as the terminal electron acceptor instead of oxygen [24]. The prevalence of NDH-1 components in the list of candidate genes, along with the presence of *rqlI* and *arcB*, highlights the importance of microaerobic metabolism in this Tn-seq experiment.

To determine in an unbiased fashion if common pathways or functions were enriched in any of these candidate gene groups, gene ontology (GO) analysis was performed using Protein Analysis Through Evolutionary Relationships (PANTHER) [25]. As a control for the GO analysis, a gene subset was also analyzed which only included

genes that were not detected in the input sample, likely due to their essential nature. This essential gene subset was found to be enriched for the translation and biosynthetic process GO categories, both of which are made up of genes important for basic processes such as translation, transcription, and replication (Fig 4.1C). No GO enrichments were observed in the glucose candidate gene subset. However, the fatty acid beta-oxidation GO category was enriched in the mucus candidate gene subset (Fig 4.1C), indicating that genes related to beta-oxidation are found in the mucus condition more often than would be expected. These genes include *fadL*, *fadE*, *fadB*, and *fadJ*, all of which have been shown to be important for the pathway through which long chain fatty acids are transported into the cell and metabolized via beta-oxidation [26, 27]. The result of beta-oxidation is several acetyl-CoA molecules, which can then enter into the citric acid cycle. These data indicate that the bacteria metabolized long chain fatty acids during the Tn-seq mucus experiment.

Many candidate genes are plasmid-encoded

There was a striking predominance of candidate genes in both the mucus and glucose conditions that are likely located on a large plasmid that is very similar to the pUTI89 plasmid carried by the cystitis isolate UTI89 [28]. There are 5 F11 contigs that closely align to the pUTI89 plasmid at the nucleotide level, with relatively small gaps between contigs (Fig 4.S1A). Several genes located on these 5 contigs met the requirements of candidate genes in both mucus and glucose. Expanding the analysis to all insertion sites found on pUTI89 demonstrated that transposon insertions in the plasmid led to reduced fitness in general (Fig 4.S1B). The median fitness score of plasmid localized insertion sites was -1.46 in the mucus broth, in comparison with 0.04 for chromosomal sites. Similar median fitness scores were observed for the glucose condition, with -1.14 for plasmid sites and 0.05 for chromosomal sites. These data

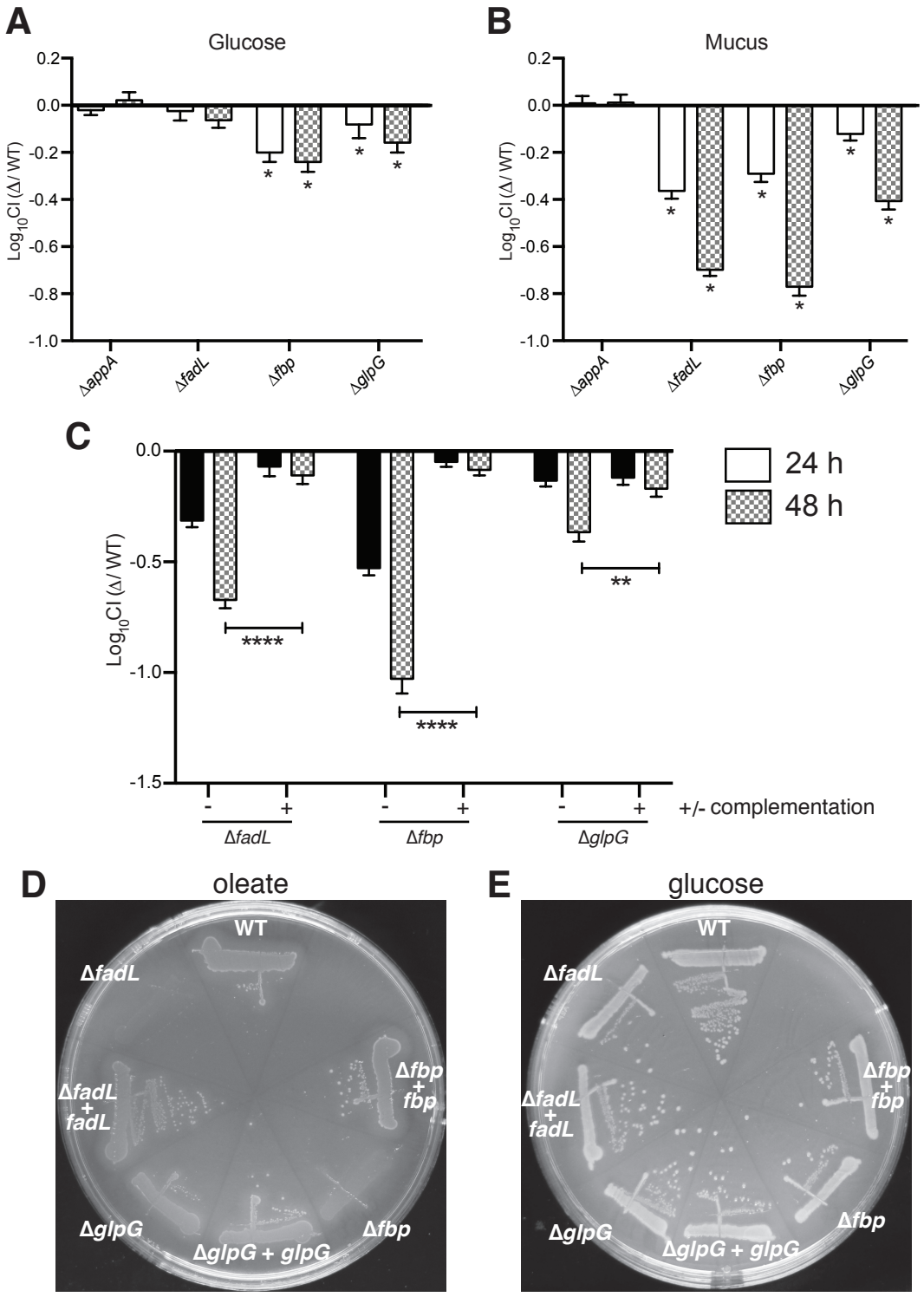
suggest that carriage of a fully-functional pUTI89 plasmid contributed to fitness in the Tn-seq experiment.

The fadL, fbp, and glpG genes are important for growth in mucus

In order to study genes important for growth in mucus, various genes were chosen from the mucus-specific candidate list, including *appA*, *fadL*, *fbp*, and *glpG*. Each of these genes was knocked out and retested in competition with the WT strain in the mucus and glucose growth assay. AppA is a periplasmic phosphatase whose expression is increased in anaerobic conditions [29]. The *fadL* gene encodes the outer membrane localized transporter of long chain fatty acids [30]. Fbp mediates the rate-limiting step of gluconeogenesis, which is important for several biosynthetic pathways when glucose is absent [31]. Finally, *glpG* is a part of the *glpEGR* operon and encodes a membrane localized protease whose substrate is unknown [32]. When each of these mutants was competed against WT in the *in vitro* growth assay, the Δ*appA* strain did not exhibit a growth defect, whereas the others did (Figs 4.2A-B). The Δ*fadL* mutant demonstrated the most striking defect as it had a competitive index (CI) near -0.7 (roughly a 5-fold decrease in fitness) at 48 h in mucus, but a CI that did not differ significantly from 0 in the glucose condition. The Δ*fbp* and Δ*glpG* mutants grew worse than WT in both the glucose and mucus conditions, but with a more pronounced defect in the mucus media (Figs 4.2A-B). The Δ*fadL*, Δ*fbp*, and Δ*glpG* mutants grew similar to WT when their respective genes or operons were returned to them on a plasmid (Fig 4.2C), further solidifying the role of these genes in mucus broth growth.

Disruptions to the beta-oxidation pathway lead to reduced growth on minimal media plates supplemented with the long chain fatty acid oleate [30]. As expected, the Δ*fadL* mutant grew poorly on oleate, but was rescued when the *fadL* gene was included on a plasmid (Fig 4.2D). Interestingly, the Δ*fbp* strain followed a similar pattern as the

Figure 4.2. Mucus-specific candidate genes are important for growth on the LCFA oleate. Four genes were chosen from the Tn-seq mucus candidate gene list, knocked out, and tested in various assays. (A) Each of the four mutants was competed against WT during microaerobic growth in M9-glucose media. Each mixture was grown for 24 h, then subcultured 1:100 and grown for another 24 h, with titering at both 24 and 48 h. The Log_{10} competitive index (CI) was calculated by dividing the mutant titers by the WT titers. * indicates that the CI is significantly different from 0 by the one-sample t-test, $P \leq 0.05$. (B) Bacteria were grown and titered as in A, except in M9-mucus media. * indicates that the CI is significantly different from 0 by the one-sample t-test, $P \leq 0.05$. (C) Those mutants that exhibited a defect in the M9-mucus media in B were grown as in B in competition against the WT strain carrying an empty vector plasmid (pCWR40). The mutants carried either the empty vector plasmid or a complementation plasmid expressing the gene or operon under their native promoters. The Δ*fadL* mutant was complemented by pCWR37, the Δ*fbp* mutant was complemented by pCWR38, and the Δ*glpG* mutant was complemented by pCWR39 which expresses the entire *glpEGR* operon. ** $P \leq 0.01$, **** $P \leq 0.0001$ by unpaired Student's t-test. Experiments in A, B, and C were performed in triplicate, three times. (D) The Δ*fadL*, Δ*fbp*, and Δ*glpG* mutants carrying either an empty vector control or a complementation plasmid as in C were streaked onto M9-oleate plates.


A
Glucose
Log10CI (Δ/ WT)
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
ΔappA
ΔfadL
Δfbp
ΔglpG
B
Mucus
C
24 h
48 h
****
****
**
+/- complementation
ΔfadL
Δfbp
ΔglpG
D
oleate
WT
ΔfadL
ΔfadL + fadL
Δfbp + fbp
ΔglpG
ΔglpG + glpG
Δfbp
E
glucose

Δ*fadL* mutant on the oleate plates, whereas the Δ*glpG* strain exhibited a subtler growth defect (Fig 4.2D). Neither Fbp nor GlpG have a previously described connection to beta-oxidation. Importantly, all of the strains grew similarly on a plate supplemented with glucose (Fig 4.2E).

GlpG is required for gut colonization, whereas FadL and Fbp are not

As the mucus layer is thought to be a source of nutrients for *E. coli* during gut colonization, and since the Δ*fadL*, Δ*fbp*, and Δ*glpG* mutants exhibited a growth defect during mucus metabolism *in vitro*, we tested their ability to colonize the mouse gut. Each mutant was competed 1:1 against WT bacteria in mouse gut colonization. When either the Δ*fadL* or Δ*fbp* mutants were competed against WT, the median competitive index was relatively close to 0 throughout the experiment, being 0.4 and -0.29 at day 14, respectively (Figs 4.3A-B). In contrast, when the Δ*glpG* mutant was competed against WT, the median competitive index dipped to -1.96 by day 14, which indicates 90-fold fewer mutant bacterial numbers in comparison to WT (Fig 4.3C). These data demonstrate that a disruption in *glpG* reduces bacterial fitness in the gut, whereas a mutation in *fadL* or *fbp* has no effect.

Mutation of glpG has polar effects on the downstream gene glpR

The *glpG* gene is the second gene of a three gene operon that includes *glpE* and *glpR* (Fig 4.4A). GlpE is a thiosulfate sulfurtransferase that is capable of transferring sulfur from thiosulfate to a recipient compound such as thioredoxin, but its biological role is not known [33]. GlpR is a transcriptional repressor of several genes that are involved with glycerol degradation [34]. There are multiple promoters that drive expression of the genes in the *glpEGR* operon, including two promoters inside of *glpG* that drive transcription of *glpR* [35]. It was possible that transposon insertions into *glpG* or replacement of *glpG* with a resistance marker would lead to polar effects on *glpR* and

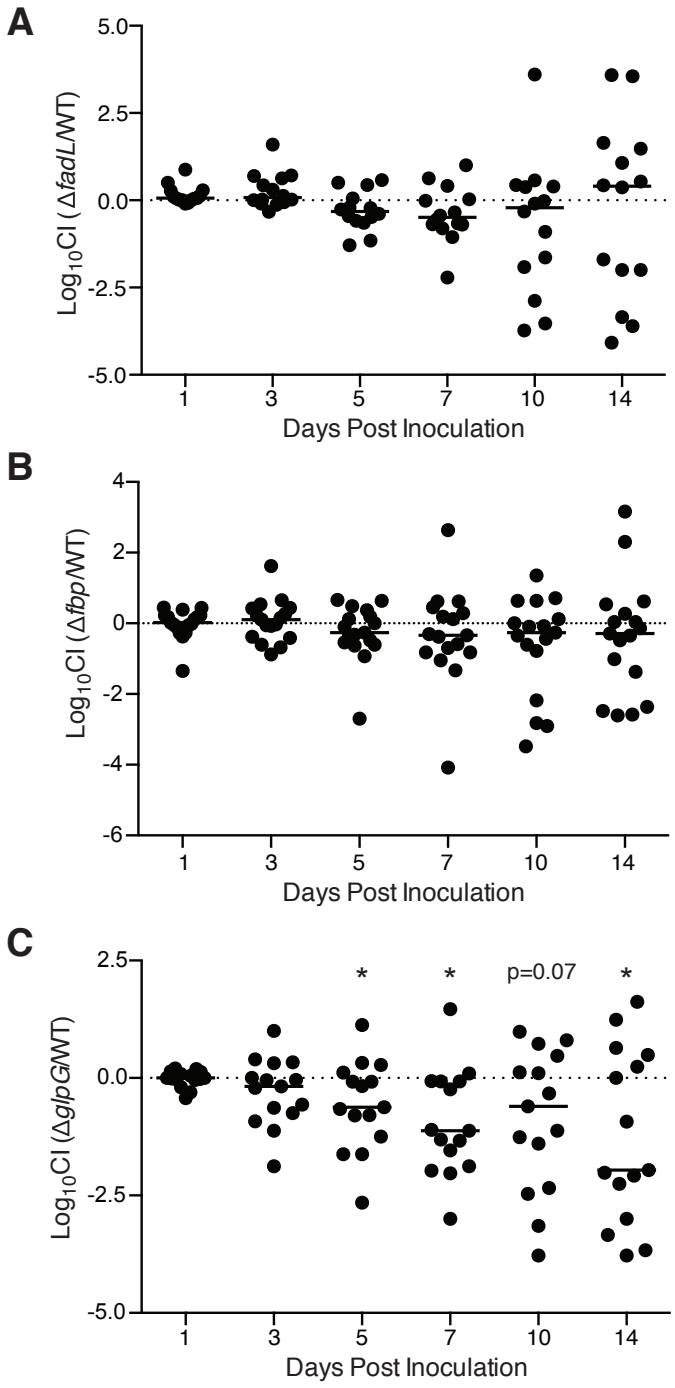


Figure 4.3. GlpG contributes to fitness during mouse gut colonization. WT and mutant bacteria were mixed 1:1 and were gavaged into mice at a dose of 10^9 CFUs. At various time points post inoculation, feces were collected, homogenized, and serially diluted onto LB containing either kanamycin to count the WT strain or chloramphenicol to count the mutant strain. For each mouse at each time point, a competitive index was calculated as the Log_{10}(mutant CFUs/WT CFUs). * indicates $P \leq 0.05$ using one sample t test with hypothetical mean set to 0. n=14-17 mice, utilized over three experiments. (A) WT vs Δ*fadL*. (B) WT vs Δ*fbp*. (C) WT vs Δ*glpG*.

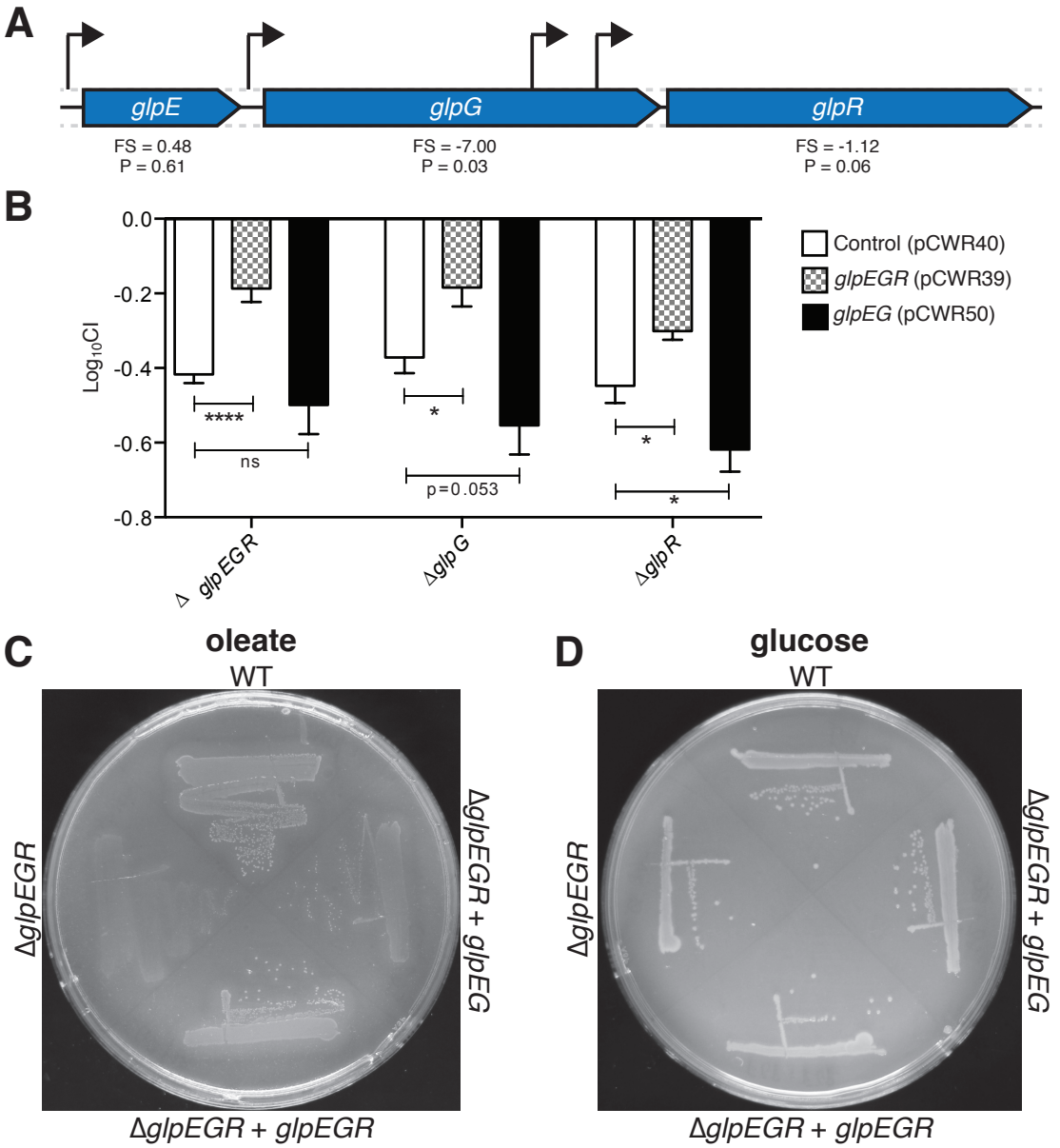


Figure 4.4. The Δ*glpG* mutant exhibits a defect in the downstream gene *glpR*. (A) A schematic of the *glpEGR* operon, with each gene indicated as a blue box. Promoter locations are indicated by black arrows above the operon. Below the operon are listed the fitness scores ("FS") and P values ("P") for each gene in the mucus growth condition in the Tn-seq experiment. (B) Each *glp* mutant carrying the empty vector control, a plasmid expressing *glpEGR*, or a plasmid expressing *glpEG*, was grown in competition with WT carrying the control plasmid. The bacteria were grown in M9-mucus media for 24 h microaerobically, and then subcultured into fresh media and grown for an additional 24 h. After 48 h of growth, each culture was titered to enumerate WT and mutant levels. * $P \leq 0.05$, **** $P \leq 0.0001$ by unpaired Student's t-test. The experiment was performed in triplicate, three times. (C) The WT and Δ*glpEGR* mutant strain carrying various expression constructs was grown on minimal media agar containing oleate or (D) glucose as the sole source of carbon.

disrupt its function. To test this possibility, the Δ*glpEGR*, Δ*glpG*, and Δ*glpR* mutants were transformed with a plasmid encoding either the entire *glpEGR* operon, or simply *glpEG*. Each of the mutants was then competed against the WT strain in a mucus growth assay (Fig 4.4B). At 24 h, all of the mutants carrying the empty vector control exhibited a similar growth defect, with a CI of about −0.4. With the addition of the *glpEGR* expression plasmid, the mutants' fitness improved. In contrast, the *glpEG* expression plasmid appeared to worsen the *glp* mutants' defect. This pattern of growth was mirrored on oleate plates for the Δ*glpEGR* strain (Fig 4.4C). Whereas the Δ*glpEGR* mutant carrying either the control plasmid or the *glpEG* expression plasmid grew more slowly than the WT control strain, mutant growth was rescued with addition of the *glpEGR* plasmid. In contrast, all of the strains grew equally well on glucose media (Fig 4.4D). Altogether, these data suggest that the reduced fitness of the various *glp* mutants is due to a lack of *glpR*.

The glp mutants are rescued during growth on oleate with the addition of glycerol-3-phosphate

As GlpR is a transcriptional repressor, its absence leads to hyperactivation of the glyercol degradation pathway, which can deplete intermediates such as glycerol-3-phosphate (G3P). G3P can be further modified in this pathway to dihydroxyacetone phosphate (DHAP) which can then be utilized in glycolysis or gluconeogenesis [36]. Additionally, G3P is an important starting substrate for membrane biogenesis [37]. Therefore, the depletion of G3P could lead to multiple detrimental downstream effects in a Δ*glpR* mutant. Indeed, it was previously observed that a Δ*glpR* mutant defect could be rescued by the addition of exogenous G3P [38].

To test whether the defect during growth on oleate could be rescued by the addition of G3P, the Δ*glpR*, Δ*glpG*, and Δ*glpEGR* mutants were grown on various

minimal medium plates. Nothing grew on the control plate that was supplemented with G3P alone (Fig 4.5A). As observed earlier, the *glp* mutants grew slower than WT on the oleate-supplemented plates (Fig 4.5B). When G3P was added in combination with the oleate, the mutants grew similarly to WT (Fig 4.5C), indicating that the mutants' growth defect could be rescued with exogenous G3P. These data indicate that the *glp* mutants may experience G3P depletion due to hyperactivation of the glycerol degradation pathway.

Discussion

To better understand how ExPEC colonizes the gut, we performed a Tn-seq experiment with the aim of defining the genes important for metabolism of intestinal mucus. Mucus is thought to be a source of nutrients for ExPEC in the gut environment given the association of commensal *E. coli* with the mucus layer in the streptomycin-treated mouse model [14, 15] and the observation that for a given strain, there is often a correlation between *in vitro* growth in mucus and fitness *in vivo* [15, 16]. By comparing the growth of individual mutants in our F11 transposon mutant library in M9-mucus with that of M9-glucose, we aimed to focus on mucus-specific factors, as opposed to general growth factors. Based on various parameters, we created a list of 112 mucus-specific candidate genes (Fig 4.1B). Four of these genes were retested in the same conditions as the Tn-seq experiment, and three of the mutants (Δ*fadL,* Δ*fbp*, and Δ*glpG*) recapitulated the Tn-seq data (Fig 4.2B). We found that the fatty acid degradation pathway was significantly enriched in the mucus specific gene set (Fig 4.1C). This finding was extended by the observation that the genes that we verified to have growth defects in mucus broth also exhibited a growth defect when oleic acid was the only carbon source (Fig 4.2D), suggesting that there may be several genes on the mucus candidate gene list that directly or indirectly affect the fatty acid degradation pathway, although they are not

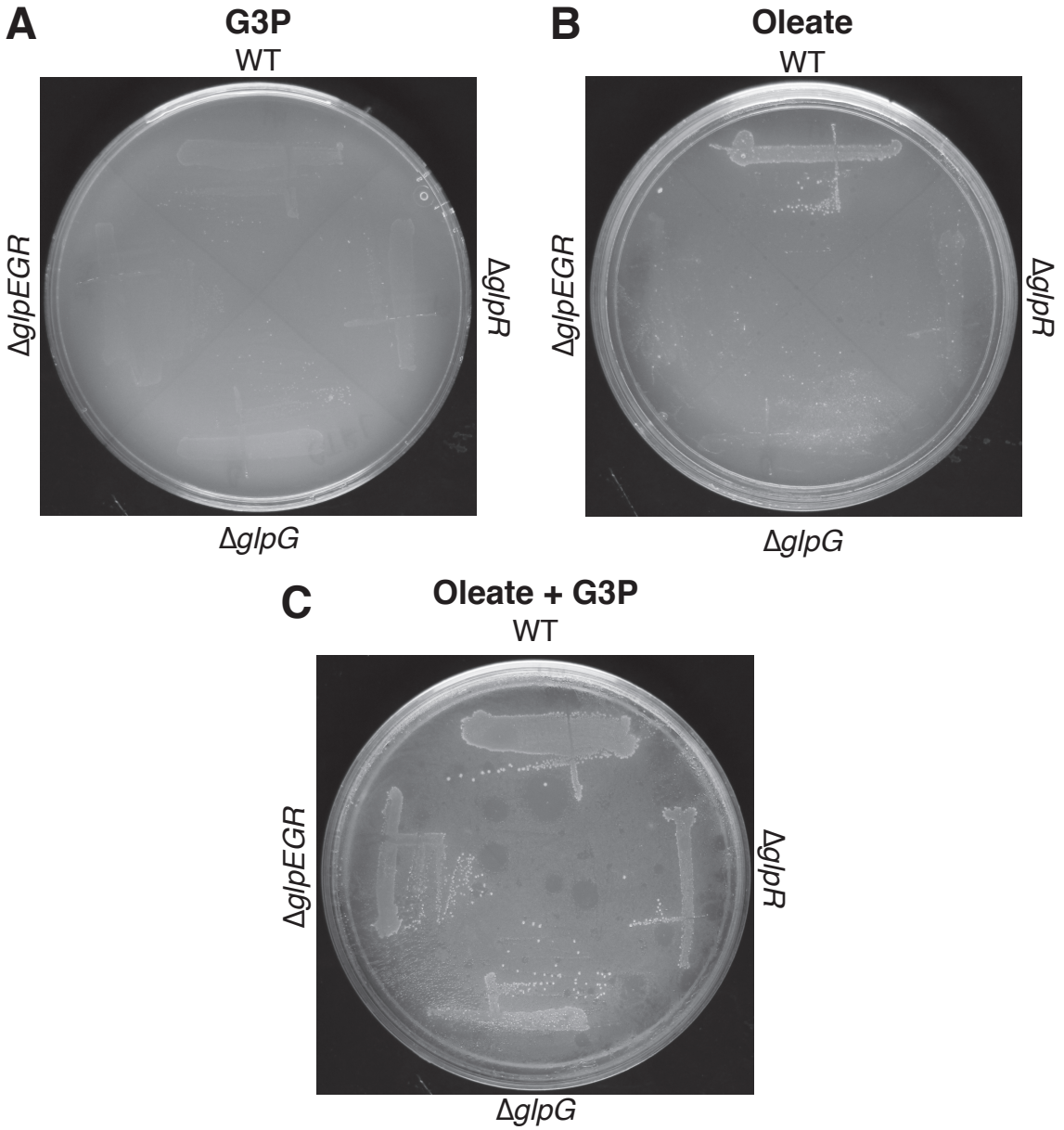


Figure 4.5. Exogenous glycerol-3-phosphate rescues the *glp* mutants' growth defects. The WT, Δ*glpR*, Δ*glpG*, and Δ*glpEGR* strains were streaked onto minimal medium plates supplemented with (A) glycerol-3-phosphate, (B) oleate, or (C) both glycerol-3-phosphate and oleate.

explicitly a part of that GO category.

It is interesting to note that sugar utilization pathways were not prominent in the mucus candidate gene list, despite the mucin protein being heavily glycosylated. This may be due to the lack of the microbiota in the Tn-seq experiment, including bacteria that would be capable of releasing sugars from mucin [13]. However, the fact that lipids are utilized in our system is not surprising. A previous study found that *E. coli* (including an ExPEC isolate) and *Salmonella* strains were capable of growing on lipids extracted from mucus, although the lipid fraction did not support growth to the same extent as whole mucus [39]. Additionally, a microarray of MG1655 genes expressed *in vitro* during growth on mouse cecal mucus demonstrated induction of genes important for utilization of the lipid phosphatidylethanolamine [18]. These studies support the notion that lipids are an edible portion of the mucus layer for bacteria.

Of the three mutants with a fitness defect *in vitro* in mucus broth, only the Δ*glpG* mutant had a defect during mouse gut colonization. Our results demonstrate that mutants defective in mucus growth *in vitro* are not always deficient during mouse gut colonization. This idea is supported by evidence that genes that are transcriptionally induced during growth on mucus *in vitro* are not necessarily required for growth *in vivo* [18]. The likely reason for this imperfect correlation is that there are several niches found within the gut, both physically and nutritionally, and some of these could support colonization in a manner independent of mucus-derived nutrients. Evidence for multiple nutritional niches during *E. coli* gut colonization comes from studies determining which *E. coli* strains can simultaneously colonize the gut [40, 41]. The extent to which two *E. coli* strains can coexist in the gut is dependent on the overlap between their respective metabolic capabilities. If the contributions of the microbiota are also considered, it is clear that there are multiple ways to obtain nutrients in the gut depending on the context, and scavenging mucus-derived products is likely only one strategy.

Why the Δ*glpG* mutant has a defect *in vivo* is not clear, although we believe it to be attributable to the polar effects that a *glpG* mutation has on *glpR* (Fig 4.4). Polar effects are assumed based on the observations that the Δ*glpG* and Δ*glpR* mutants phenocopy in the mucus broth experiment, and that they are both rescued by a plasmid carrying *glpEGR*, but not a plasmid expressing *glpEG* (Fig 4.4B). As GlpR is a transcriptional repressor of the glycerol degradation pathway, a Δ*glpR* mutant has a hyperactive glycerol degradation pathway which can act as a double-edged sword. On one hand, a Δ*glpR* mutation allows for improved growth on glycerol and is thought to be the reason that many K-12 strains have acquired spontaneous *glpR* mutations [42]. However, increased glycerol degradation can deplete glycerol-3-phosphate, an intermediate in the pathway that is also important for membrane biogenesis [38]. Therefore, as the Δ*glpG* mutant exhibits a defect in mouse gut colonization, it is clear that ExPEC needs to carefully modulate the levels of glycerol degradation *in vivo*. A BLAST search found that the *glpEGR* genes appear to be restricted to the Enterobacteriaceae family, suggesting that these findings may be applicable to other genera including *Klebsiella*, *Shigella*, and *Salmonella*.

In this work we set out to find genes required for ExPEC gut colonization, as the gut is believed to be the primary ExPEC niche. Using data from an *in vitro* Tn-seq experiment, we endeavored to find mutants with gut colonization defects. Through these efforts it was found that proper regulation of the glycerol degradation pathway is required for persistence of ExPEC in the intestine. Future work to delineate gut colonization by ExPEC will be invaluable for obtaining a broader view of the ExPEC life cycle, and may possibly contribute to development of novel therapeutics.

Materials and Methods

Creation of bacterial strains and plasmids

The cystitis isolate F11 was genetically altered using strains carrying pKM208, which facilitates creation of knockouts and knockins via lambda-red recombination [43]. Knockout and knockin constructs were created by PCR using either pKD3 or pKD4 as template to make a chloramphenicol or kanamycin resistance cassette, respectively, along with ~40 bps of flanking overhang with homology to the genomic insertion site.

All plasmids utilized in this study are derived from pACYC177 [44]. All of the expression vectors were created by traditional cloning techniques, with insertion of the PCR product at the BamHI and NheI sites, which removes the kanamycin resistance cassette. The control plasmid pCWR40 was created by digestion of pACYC177 with BamHI and NheI, followed by a reaction with the Klenow fragment to create blunt ends prior to re-ligation.

Media

While creating knockout, knockin, and complemented strains, bacteria were grown in Luria-Bertani (LB) broth. All growth in petri dishes was done using LB agar supplemented with chloramphenicol (20 μg/mL), kanamycin (50 μg/mL), or ampicillin (100 μg/mL). When growing bacteria for mouse experiments, a modified M9 medium was utilized. It contains $MgSO_4 \bullet 7H_2O$ (1 mM), $CaCl_2 \bullet 2H_2O$ (0.1 mM), D-(+)-Glucose (0.1%), Nicotinic Acid (0.00125%), Thiamine HCl (0.00165%), Casamino Acids (0.2%), Na_2HPO_4 (6g/L), KH_2PO_4 (3 g/L), NH_4Cl (1 g/L), and NaCl (0.5 g/L), all dissolved in water. For *in vitro* growth experiments in “M9-glucose” or “M9-mucus”, the above recipe was used, except that Casamino Acids were excluded. Additionally, for the M9-mucus media, 0.5% mucus (Sigma M1778) was used in place of the glucose. To make M9-mucus, the mucus was dissolved in water by autoclaving for 20 min at 121° C.

Mucus and glucose Tn-seq experiment

Three independently-derived transposon mutant pools that were used previously [20] were utilized in this experiment. Briefly, these pools were created by conjugative transfer to F11 of the pSAM-Ec plasmid, which contains a himar1C9 transposase and a mariner transposon with a kanamycin resistance cassette. Transposon mutants were selected by plating on LB supplemented with kanamycin, and further analysis indicated few ampicillin resistant bacteria, indicating that whole plasmid insertion occurred infrequently.

Each of the transposon mutant pools was expanded in LB broth, and then subcultured 1:100 into M9-mucus or M9-glucose, and grown at 37° C for 24 h shaking at 180 rpm in a jar containing an AnaeroPack-MicroAero sachet. Each culture was then subcultured into fresh media and grown for another 24 h. After 48 h of growth, cells were collected for DNA purification.

DNA was prepared for sequencing via HTML-PCR as described [45], with some modifications. DNA from the “input” (LB) and “output” (M9-glucose and M9-mucus) samples was isolated using a Wizard Genomic DNA Purification Kit (Promega). Thirty μg of genomic DNA were diluted in 50 μl of water and sonicated using a Diagenode Sonicator, using 30 cycles of 30 seconds ON, 60 seconds OFF. This treatment led to sheared DNA being mostly less than 600 bps. The DNA was purified and concentrated using the GeneJet PCR Purificaiton Kit (Fermentas), which removes DNA pieces below the size of 100 bps. Poly-dC tails were added to the 3’ ends of the sheared DNA pieces by incubation of 1.5 μg of DNA with Terminal Deoxynucleotidyl Transferase (TdT from NEB) at 37° C for 1 h, followed by heat inactivation at 75° C for 20 minutes. The TdT reaction was purified using a DTR Gel Filtration Cartridge (Performa) following the directions provided by the manufacturer. Two PCR reactions were performed in succession using the Easy A Cloning Enzyme. The first reaction contained primers olj376

and tn_rev1-2, while the second contained primers tn_rev2-2 and the appropriate indexing primer. The reactions were purified and were sent for sequencing by the Tufts University Genomics Core Facility using the "pSAM-Ec Illumina seq" primer. Sequencing was done on the HiSeq 2500, high output v4, with single-end reads produced over 50 cycles.

Bioinformatic analysis

Reads were initially processed using a custom Python script to remove adaptor sequences, trim low quality bases from the 3' end of the read (minimum phred score of 20 required), trim strings of C nucleotides that may have originated from the TdT reaction from the 3' end, as well as to filter out reads with low quality 5' ends (minimum phred score of 20 required). After trimming, any read with a length less than 28 nucleotides was removed. On average, these requirements led to ~3% of reads being removed.

The trimmed reads were then aligned to the F11 genome using Bowtie2, and the resulting SAM files were sorted using SAMtools. Custom Python scripts were used to analyze the sorted SAM files, including identification of transposon insertion locations, and quantification of the number of reads for each insertion location. Lower quality alignments were filtered out by requiring that each alignment have a Mapq score of at least 35, no more than 2 mismatches, and that it not align equally well to more than one location on the genome. Additionally, alignments were required to correspond to a TA dinucleotide, the preferred insertion location of the mariner transposon. These thresholds resulted in an additional ~3% of the reads being filtered at this step. During this process, forward and reverse insertion events that occurred at the same location were pooled together and considered as one insertion site.

In order to calculate fitness scores for individual transposon insertion sites, a

table was created for each input-output sample pair that included the read counts for only those insertion sites that were detected in both samples. Insertion sites were required to have at least 10 reads in the input sample to be included. Transposon mutants that were completely lost from the output sample were added to the table if the number of reads at that insertion site in the input sample was very high, within the top 1% of input sites. In this case, the output read count was set to 1, the limit of detection. This practice allowed for analysis of mutants with very low fitness in the growth conditions. The input-output pair tables were then normalized using geometric mean normalization, similar to the normalization strategy utilized by DESeq2 [46]. The normalized reads were then used to calculate a fitness score for each insertion site in each sample, which was calculated as Log_2(output sample reads/input sample reads). The fitness scores for all of the insertion sites in a given gene across the three replicates were used to calculate a median fitness score, which was designated the fitness score for that gene. A P value was calculated using a Wilcoxon Signed Rank Test with the hypothetical median set to 0.

A list of candidate genes was created with the genes that met certain criteria, which included a fitness score of -2.0 or lower (i.e. at least a 4-fold reduction in fitness from input to output) and a P value of 0.05 or lower. The candidate gene lists were analyzed using gene ontology via PANTHER [25].

Mouse gut colonization

Mice were handled in accordance with protocols approved by the Institutional Animal Care and Use Committee at the University of Utah (Protocol number 15-12015), following US federal guidelines indicated by the Office of Laboratory Animal Welfare (OLAW) and described in the Guide for the Care and Use of Laboratory Animals, 8th Edition.

For mouse gut colonization experiments, female Balb/c mice, aged 7-8 weeks, were intragastrically gavaged with 50 µl of PBS containing 10^9 CFUs of bacteria. The mice were housed 3-5 per cage, and were allowed to eat (Teklad Global Soy Protein-Free Extruded, irradiated, 2920x) and drink *ad libitum*. Feces were collected at various time points, homogenized in 1 mL of 0.7% NaCl, and briefly centrifuged at a low speed to pellet any insoluble debris. Supernatants were then serially diluted and plated onto LB plates containing either chloramphenicol (20 µg/mL) or kanamycin (50 µg/mL) to select for the relevant strains. Fecal samples were also collected and plated prior to gavage to ensure that the microbiota did not contain any bacteria that could be cultured on chloramphenicol or kanamycin.

The bacteria for the mouse gut experiments were grown statically in 20 mL modified M9 in a 250 mL flask for 24 h at 37° C. For noncompetitive experiments, 12 mL of culture was centrifuged at 8,000 r.c.f. for 10 minutes, and the pellet was washed and resuspended in 0.5 mL of PBS. For competitive experiments, 6 mL of culture for each strain were mixed prior to centrifugation.

Growth on oleate plates

To make M9-oleate plates, the reagents for the modified M9 media as outlined in the "Media" section were mixed, with the exception of casamino acids and glucose. Triton X-100 (0.6%) and oleate (0.1%) were also added, and everything was allowed to rock at room temperature for 20 minutes to facilitate mixture of the oleate with the Triton X. Molten agar was added to the mixture and the plate was poured. Control plates were made by adding glucose (0.5%) or glycerol-3-phosphate (1 mM) instead of oleate. Once the plates were made, 1 µl of overnight culture from each strain was streaked out and the plates were incubated at 37° C.

Statistical analysis

Statistical analysis for the Tn-seq experiment was performed using the NumPy package. All other analysis was done in GraphPad Prism. The specific statistical test used in each case is indicated in the figure legend.

References

1. Barber AE, Norton JP, Spivak AM, Mulvey MA. Urinary tract infections: current and emerging management strategies. Clin Infect Dis. 2013;57(5):719-24. doi: 10.1093/cid/cit284. PubMed PMID: 23645845; PubMed Central PMCID: PMCPMC3739462.

2. Foxman B, Barlow R, D'Arcy H, Gillespie B, Sobel JD. Urinary tract infection: self-reported incidence and associated costs. Ann Epidemiol. 2000;10(8):509-15. PubMed PMID: 11118930.

3. Hooton TM. Recurrent urinary tract infection in women. Int J Antimicrob Agents. 2001;17(4):259-68. PubMed PMID: 11295405.

4. Foxman B. The epidemiology of urinary tract infection. Nat Rev Urol. 2010;7(12):653-60. doi: 10.1038/nrurol.2010.190. PubMed PMID: 21139641.

5. Simonsen KA, Anderson-Berry AL, Delair SF, Davies HD. Early-onset neonatal sepsis. Clin Microbiol Rev. 2014;27(1):21-47. doi: 10.1128/CMR.00031-13. PubMed PMID: 24396135; PubMed Central PMCID: PMCPMC3910904.

6. Nicolas-Chanoine MH, Bertrand X, Madec JY. Escherichia coli ST131, an intriguing clonal group. Clin Microbiol Rev. 2014;27(3):543-74. doi: 10.1128/CMR.00125-13. PubMed PMID: 24982321; PubMed Central PMCID: PMCPMC4135899.

7. Moreno E, Andreu A, Pigrau C, Kuskowski MA, Johnson JR, Prats G. Relationship between Escherichia coli strains causing acute cystitis in women and the fecal E. coli population of the host. J Clin Microbiol. 2008;46(8):2529-34. doi: 10.1128/JCM.00813-08. PubMed PMID: 18495863; PubMed Central PMCID: PMCPMC2519474.

8. Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME, Hultgren SJ, et al. Genomic diversity and fitness of E. coli strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Sci Transl Med. 2013;5(184):184ra60. doi: 10.1126/scitranslmed.3005497. PubMed PMID: 23658245; PubMed Central PMCID: PMCPMC3695744.

9. Foxman B, Manning SD, Tallman P, Bauer R, Zhang L, Koopman JS, et al. Uropathogenic Escherichia coli are more likely than commensal E. coli to be shared between heterosexual sex partners. Am J Epidemiol. 2002;156(12):1133-40. PubMed PMID: 12480658.

10. Hilali F, Ruimy R, Saulnier P, Barnabe C, Lebouguenec C, Tibayrenc M, et al.

Prevalence of virulence genes and clonality in Escherichia coli strains that cause bacteremia in cancer patients. Infect Immun. 2000;68(7):3983-9. PubMed PMID: 10858212; PubMed Central PMCID: PMCPMC101677.

11. Cortese F, Scicchitano P, Gesualdo M, Filaninno A, De Giorgi E, Schettini F, et al. Early and Late Infections in Newborns: Where Do We Stand? A Review. Pediatr Neonatol. 2015. doi: 10.1016/j.pedneo.2015.09.007. PubMed PMID: 26750406.

12. Johansson ME, Phillipson M, Petersson J, Velcich A, Holm L, Hansson GC. The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc Natl Acad Sci U S A. 2008;105(39):15064-9. doi: 10.1073/pnas.0803124105. PubMed PMID: 18806221; PubMed Central PMCID: PMCPMC2567493.

13. Conway T, Cohen PS. Commensal and Pathogenic Escherichia coli Metabolism in the Gut. Microbiol Spectr. 2015;3(3). doi: 10.1128/microbiolspec.MBP-0006-2014. PubMed PMID: 26185077; PubMed Central PMCID: PMCPMC4510460.

14. Poulsen LK, Lan F, Kristensen CS, Hobolth P, Molin S, Krogfelt KA. Spatial distribution of Escherichia coli in the mouse large intestine inferred from rRNA in situ hybridization. Infect Immun. 1994;62(11):5191-4. PubMed PMID: 7927805; PubMed Central PMCID: PMCPMC303247.

15. Moller AK, Leatham MP, Conway T, Nuijten PJ, de Haan LA, Krogfelt KA, et al. An Escherichia coli MG1655 lipopolysaccharide deep-rough core mutant grows and survives in mouse cecal mucus but fails to colonize the mouse large intestine. Infect Immun. 2003;71(4):2142-52. PubMed PMID: 12654836; PubMed Central PMCID: PMCPMC152069.

16. Wadolkowski EA, Laux DC, Cohen PS. Colonization of the streptomycin-treated mouse large intestine by a human fecal Escherichia coli strain: role of growth in mucus. Infect Immun. 1988;56(5):1030-5. PubMed PMID: 3281898; PubMed Central PMCID: PMCPMC259757.

17. Sweeney NJ, Laux DC, Cohen PS. Escherichia coli F-18 and E. coli K-12 eda mutants do not colonize the streptomycin-treated mouse large intestine. Infect Immun. 1996;64(9):3504-11. PubMed PMID: 8751891; PubMed Central PMCID: PMCPMC174255.

18. Chang DE, Smalley DJ, Tucker DL, Leatham MP, Norris WE, Stevenson SJ, et al. Carbon nutrition of Escherichia coli in the mouse intestine. Proc Natl Acad Sci U S A. 2004;101(19):7427-32. doi: 10.1073/pnas.0307888101. PubMed PMID: 15123798; PubMed Central PMCID: PMCPMC409935.

19. van Opijnen T, Camilli A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol. 2013;11(7):435-42. doi: 10.1038/nrmicro3033. PubMed PMID: 23712350; PubMed Central PMCID: PMCPMC3842022.

20. Wiles TJ, Norton JP, Russell CW, Dalley BK, Fischer KF, Mulvey MA. Combining quantitative genetic footprinting and trait enrichment analysis to identify fitness determinants of a bacterial pathogen. PLoS Genet. 2013;9(8):e1003716. doi: 10.1371/journal.pgen.1003716. PubMed PMID: 23990803; PubMed Central PMCID:

PMCPMC3749937.

21. Russell CW, Mulvey MA. The Extraintestinal Pathogenic Escherichia coli Factor RqlI Constrains the Genotoxic Effects of the RecQ-Like Helicase RqlH. PLoS Pathog. 2015;11(12):e1005317. doi: 10.1371/journal.ppat.1005317. PubMed PMID: 26636713; PubMed Central PMCID: PMCPMC4670107.

22. Iuchi S, Matsuda Z, Fujiwara T, Lin EC. The arcB gene of Escherichia coli encodes a sensor-regulator protein for anaerobic repression of the arc modulon. Mol Microbiol. 1990;4(5):715-27. PubMed PMID: 2201868.

23. Alexeeva S, Hellingwerf KJ, Teixeira de Mattos MJ. Requirement of ArcA for redox regulation in Escherichia coli under microaerobic but not anaerobic or aerobic conditions. J Bacteriol. 2003;185(1):204-9. PubMed PMID: 12486057; PubMed Central PMCID: PMCPMC141817.

24. Tran QH, Bongaerts J, Vlad D, Unden G. Requirement for the proton-pumping NADH dehydrogenase I of Escherichia coli in respiration of NADH to fumarate and its bioenergetic implications. Eur J Biochem. 1997;244(1):155-60. PubMed PMID: 9063459.

25. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 2016;44(D1):D336-42. doi: 10.1093/nar/gkv1194. PubMed PMID: 26578592; PubMed Central PMCID: PMCPMC4702852.

26. Fujita Y, Matsuoka H, Hirooka K. Regulation of fatty acid metabolism in bacteria. Mol Microbiol. 2007;66(4):829-39. doi: 10.1111/j.1365-2958.2007.05947.x. PubMed PMID: 17919287.

27. Campbell JW, Morgan-Kiss RM, Cronan JE, Jr. A new Escherichia coli metabolic competency: growth on fatty acids by a novel anaerobic beta-oxidation pathway. Mol Microbiol. 2003;47(3):793-805. PubMed PMID: 12535077.

28. Cusumano CK, Hung CS, Chen SL, Hultgren SJ. Virulence plasmid harbored by uropathogenic Escherichia coli functions in acute stages of pathogenesis. Infect Immun. 2010;78(4):1457-67. doi: 10.1128/IAI.01260-09. PubMed PMID: 20123719; PubMed Central PMCID: PMCPMC2849428.

29. Dassa E, Cahu M, Desjoyaux-Cherel B, Boquet PL. The acid phosphatase with optimum pH of 2.5 of Escherichia coli. Physiological and Biochemical study. J Biol Chem. 1982;257(12):6669-76. PubMed PMID: 6282821.

30. Nunn WD, Simons RW. Transport of long-chain fatty acids by Escherichia coli: mapping and characterization of mutants in the fadL gene. Proc Natl Acad Sci U S A. 1978;75(7):3377-81. PubMed PMID: 356053; PubMed Central PMCID: PMCPMC392779.

31. Fraenkel DG, Horecker BL. Fructose-1, 6-diphosphatase and acid hexose phosphatase of Escherichia coli. J Bacteriol. 1965;90(4):837-42. PubMed PMID: 4284917; PubMed Central PMCID: PMCPMC315746.

32. Wang Y, Zhang Y, Ha Y. Crystal structure of a rhomboid family intramembrane protease. Nature. 2006;444(7116):179-80. doi: 10.1038/nature05255. PubMed PMID: 17051161.

33. Ray WK, Zeng G, Potters MB, Mansuri AM, Larson TJ. Characterization of a 12-kilodalton rhodanese encoded by glpE of Escherichia coli and its interaction with thioredoxin. J Bacteriol. 2000;182(8):2277-84. PubMed PMID: 10735872; PubMed Central PMCID: PMCPMC111278.

34. Lin EC. Glycerol dissimilation and its regulation in bacteria. Annu Rev Microbiol. 1976;30:535-78. doi: 10.1146/annurev.mi.30.100176.002535. PubMed PMID: 825019.

35. Yang B, Larson TJ. Multiple promoters are responsible for transcription of the glpEGR operon of Escherichia coli K-12. Biochim Biophys Acta. 1998;1396(1):114-26. PubMed PMID: 9524241.

36. Iuchi S, Cole ST, Lin EC. Multiple regulatory elements for the glpA operon encoding anaerobic glycerol-3-phosphate dehydrogenase and the glpD operon encoding aerobic glycerol-3-phosphate dehydrogenase in Escherichia coli: further characterization of respiratory control. J Bacteriol. 1990;172(1):179-84. PubMed PMID: 2403539; PubMed Central PMCID: PMCPMC208416.

37. Dowhan W. Molecular basis for membrane phospholipid diversity: why are there so many lipids? Annu Rev Biochem. 1997;66:199-232. doi: 10.1146/annurev.biochem.66.1.199. PubMed PMID: 9242906.

38. Flower AM. SecG function and phospholipid metabolism in Escherichia coli. J Bacteriol. 2001;183(6):2006-12. doi: 10.1128/JB.183.6.2006-2012.2001. PubMed PMID: 11222599; PubMed Central PMCID: PMCPMC95096.

39. Krivan HC, Franklin DP, Wang W, Laux DC, Cohen PS. Phosphatidylserine found in intestinal mucus serves as a sole source of carbon and nitrogen for salmonellae and Escherichia coli. Infect Immun. 1992;60(9):3943-6. PubMed PMID: 1500206; PubMed Central PMCID: PMCPMC257417.

40. Meador JP, Caldwell ME, Cohen PS, Conway T. Escherichia coli pathotypes occupy distinct niches in the mouse intestine. Infect Immun. 2014;82(5):1931-8. doi: 10.1128/IAI.01435-13. PubMed PMID: 24566621; PubMed Central PMCID: PMCPMC3993424.

41. Miranda RL, Conway T, Leatham MP, Chang DE, Norris WE, Allen JH, et al. Glycolytic and gluconeogenic growth of Escherichia coli O157:H7 (EDL933) and E. coli K-12 (MG1655) in the mouse intestine. Infect Immun. 2004;72(3):1666-76. PubMed PMID: 14977974; PubMed Central PMCID: PMCPMC355998.

42. Freddolino PL, Amini S, Tavazoie S. Newly identified genetic variations in common Escherichia coli MG1655 stock cultures. J Bacteriol. 2012;194(2):303-6. doi: 10.1128/JB.06087-11. PubMed PMID: 22081388; PubMed Central PMCID: PMCPMC3256642.

43. Murphy KC, Campellone KG. Lambda Red-mediated recombinogenic engineering of

enterohemorrhagic and enteropathogenic E. coli. BMC Mol Biol. 2003;4:11. doi: 10.1186/1471-2199-4-11. PubMed PMID: 14672541; PubMed Central PMCID: PMCPMC317293.

44. Chang AC, Cohen SN. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J Bacteriol. 1978;134(3):1141-56. PubMed PMID: 149110; PubMed Central PMCID: PMCPMC222365.

45. Lazinski DW, Camilli A. Homopolymer tail-mediated ligation PCR: a streamlined and highly efficient method for DNA cloning and library construction. Biotechniques. 2013;54(1):25-34. doi: 10.2144/000113981. PubMed PMID: 23311318; PubMed Central PMCID: PMCPMC3605734.

46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PubMed PMID: 25516281; PubMed Central PMCID: PMCPMC4302049.

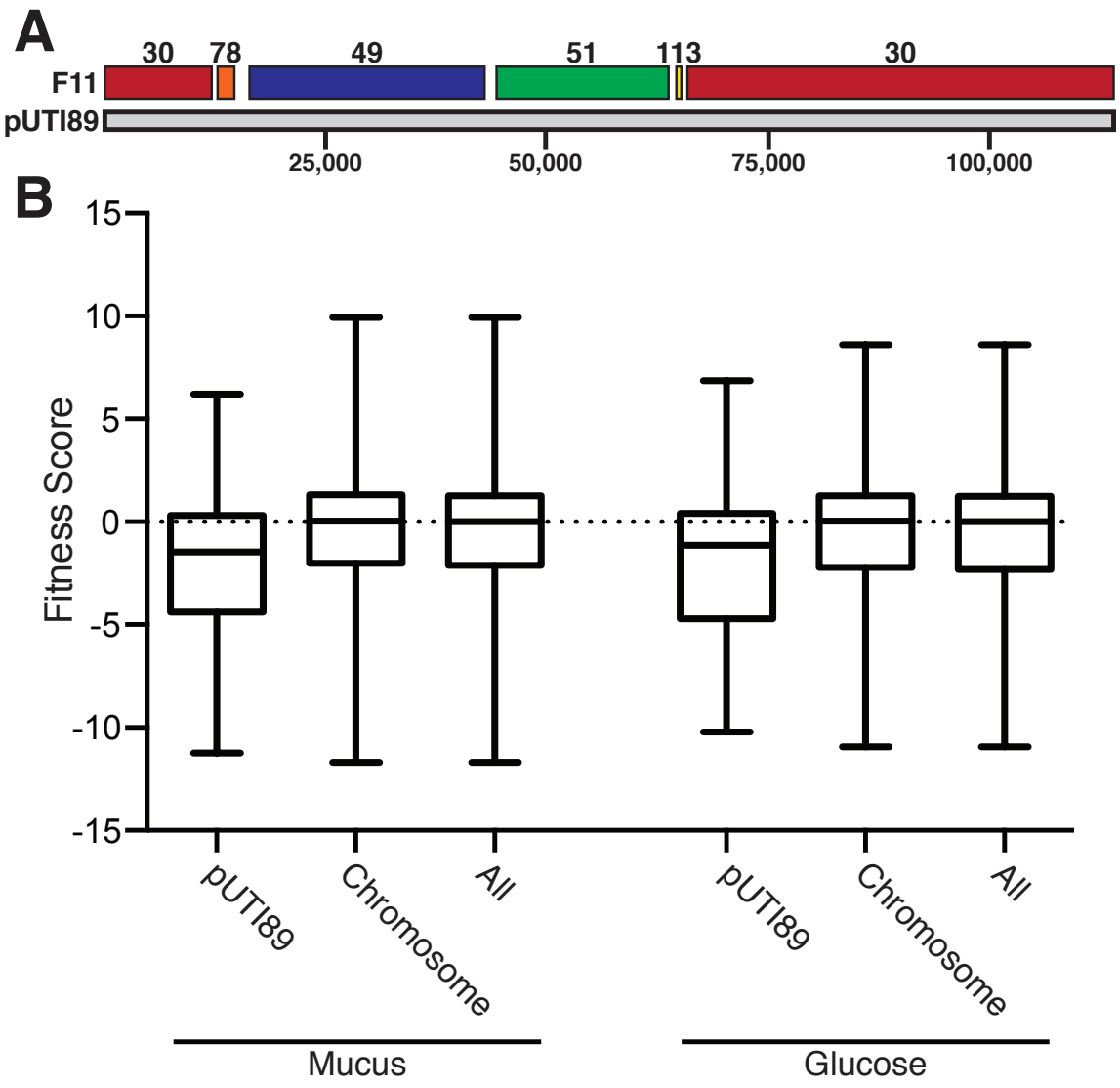


Figure 4.S1. Transposon insertions in pUTI89 reduce fitness *in vitro*. (A) A schematic of the F11 contigs that align with pUTI89. The F11 contig number is indicated at top. (B) Fitness scores for the insertion sites that are located on the F11 contigs that align with pUTI89, the sites that are on contigs that do not align with pUTI89, and all of the sites. The whiskers extend to minimum and maximum values.

CHAPTER 5

WITH THE EXCPETION OF TYPE 1 PILI, MOST EXPEC VIRULENCE FACTORS DO NOT FOLLOW A COINCIDENTAL EVOLUTIONARY PATTERN

Colin W. Russell, Alan T. Stenquist, Morgan A. Wambaugh, Mary P. Bronner, Matthew A. Mulvey

Abstract

Escherichia coli is a frequent member of the human microbiota that can exert diverse effects on the gut ecosystem. One subset of the *E. coli* species is extraintestinal pathogenic *E. coli* (ExPEC), which acts as a commensal within the gut, but induces pathology upon dissemination to other niches such as the urinary tract and bloodstream. Given that ExPEC spends the majority of its time within the gut, it is likely that it has been mostly shaped by the evolutionary pressures found in this environment, provoking the question of how its extraintestinal virulence arose. The principle of coincidental evolution, in which a factor that evolved in one niche happens to be useful in another, has been hypothesized to be applicable to the ExPEC life cycle, in that extraintestinal virulence factors in actuality arose for full gut fitness. To test this hypothesis, several mutants were made that lacked canonical ExPEC virulence factors, and their importance for gut colonization was assessed. We found that most of the factors—including Cnf1, Usp, colibactin, flagella, and the plasmid pUTI89—were dispensable for gut colonization. A couple of mutants, Δ*papG* and Δ*hlyA*, exhibited transient defects that did not affect long-term persistence. Only FimH, the adhesin for type 1 pili, was found to be important for gut colonization. Importantly, type 1 pili were important for repopulation of the gut after a course of oral antibiotics, uncovering a possible mechanism of recurrent infections. As FimH was the only virulence factor that was important for gut colonization, it is clear that coincidental evolution cannot be applied to ExPEC wholesale, but may need to be tested on a gene-by-gene basis.

Introduction

The bacterium *Escherichia coli* was first described by Theodor Escherich in 1884, and has since become a critical model organism that has been used to understand the fundamentals of molecular biology [1]. *E. coli* is able to live in a variety of locations,

including the soil, water, and the human gut—in part because it is a facultative anaerobe, meaning that it is capable of switching between aerobic and anaerobic metabolism. Although it is a prominent member of the neonatal microbiota, it is quickly overshadowed by the burgeoning anaerobic population as oxygen becomes scarce within the gut shortly after birth [2, 3]. In adults, *E. coli* is present at about 10^7 to 10^9 CFUs/g feces, a level that is 100-10,000-fold lower than the resident anaerobes [4]. Despite being a minor member of the microbiota, the estimated number of *E. coli* cells that are transmitted via fecal matter from a human being to the environment in a single day is staggering: about 10^{11} CFUs [5].

Understanding the role of *E. coli* within the microbiota is complicated by the fact that *E. coli* is a very diverse species with a wide spectrum of phenotypes [1]. Some *E. coli* strains live harmlessly in the gut, while others act as pathogens in the gastrointestinal tract, causing diarrhea and hemorrhaging [6], or have even been associated with Crohn's Disease [7]. One strain, Nissle 1917, acts as a probiotic that assuages inflammation in addition to inhibiting colonization by pathogens such as *Salmonella* [8, 9]. Finally, a group of strains known as extraintestinal pathogenic *Escherichia coli* (ExPEC) are generally commensals within the gut, but disseminate to and infect other host sites [10]. ExPEC includes uropathogenic *E. coli* (UPEC), which cause the overwhelming majority of urinary tract infections (UTIs), a bacterial infection that is especially prevalent among women, infecting about half of all women during their lifetime [11]. Other, more serious conditions can also be initiated by ExPEC, including neonatal sepsis, urosepsis, and neonatal meningitis [10].

The gut is thought to be one of the major ExPEC reservoirs that seeds extraintestinal infections. The majority of adults carry ExPEC within their gut, even in the absence of extraintestinal infection [12], indicating that the gut is capable of harboring ExPEC at a high frequency. Longitudinal studies indicate that the presence of

ExPEC within the gut is fairly stable, with the *E. coli* phylogenetic group that includes ExPEC being the most likely to reside long term when compared to other *E. coli* groups [13, 14]. Additionally, the gut ExPEC population is often difficult to clear with antibiotics, even when the urinary tract population is eradicated [15]. Together, these observations suggest that the gut is available to act as a source of ExPEC that could then travel to other host niches. More direct evidence for this notion is that the same strain can be isolated from the feces and urine of patients suffering from a UTI [16-19] indicating some sort of transit between the two niches.

The idea that ExPEC spends most of its time within the gut with only occasional dissemination to extraintestinal niches provokes the question of how ExPEC virulence has arisen. Many hypotheses have been proposed to explain how virulence evolves in pathogens in general, especially in cases where high virulence leads to cessation of pathogen transmission, such as in the event of host death [20]. In the case of ExPEC, blood infection can be thought of as a dead-end as it often leads to host expiration, especially if left untreated. Urinary tract infections are typically not as severe, allowing urine to possibly serve as a route of transmission. However, the rate of transmission between hosts via urine when compared to feces is likely low due to the extremely high rate of bacterial shedding in the feces. The hypothesis of coincidental evolution has often been evoked to explain the evolution of ExPEC virulence [21]. In general terms, coincidental evolution is when a factor evolves in one context, but happens to be useful in another context as well [20]. When the hypothesis of coincidental evolution is applied to the ExPEC life cycle, the primary niche and main evolutionary arena is the gut. The implication is that factors that promote virulence in extraintestinal niches evolved in the gut for a function possibly unrelated to virulence.

Little concrete evidence has been put forth to support or contradict coincidental evolution in the context of ExPEC infection, other than that known extraintestinal

virulence factors are often found in gut isolates [14, 21, 22]. One prediction of this hypothesis is that extraintestinal virulence factors should play a role in gut colonization, which has only been addressed by one study thus far [5]. It is clear that more experimental work needs to be done to determine to what extent coincidental evolution applies to ExPEC.

Here, we take a systematic approach to test the hypothesis of coincidental evolution as applied to ExPEC by determining the role of individual virulence factors in gut colonization. Several canonical ExPEC virulence factors were deleted and the resulting mutant was tested for its ability to colonize the mouse gastrointestinal tract. Most of these virulence factors (HlyA, Cnf1, Usp, colibactin, flagella, pUTI89, and PapG) do not appear to play a role in persistence in the gut. Only FimH, the adhesin for type 1 pili, is required for full fitness in the gut. The Δ*fimH* knockout mutant exhibits a defect in gut colonization that includes an increased frequency of clearance. This defect can be partially attributed to increased motility. We observe that the importance of FimH for gut persistence is especially prominent in the face of antibiotic treatment. Overall, these data indicate that coincidental evolution must be applied on a per-gene basis and may not apply to ExPEC as a whole. Furthermore, many of the canonical ExPEC virulence factors have no role in the gut, indicating that coincidental evolution may not be a reasonable explanation for their existence.

Results

ExPEC benignly colonizes the murine lower gastrointestinal tract in the presence of an intact microbiota

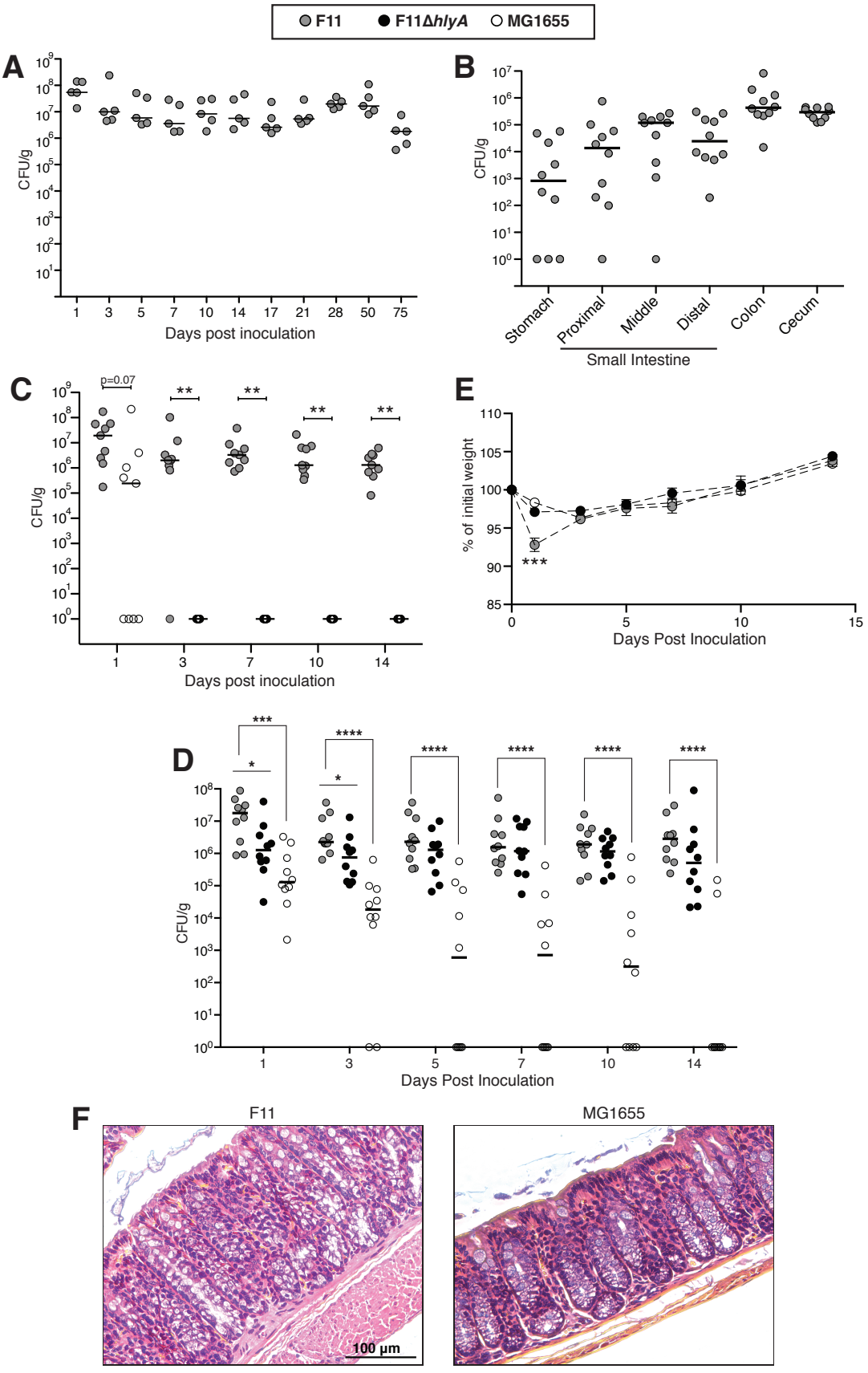
In order to establish an *in vivo* model of ExPEC gut colonization, we tested how well ExPEC could persist within the mouse gut. Specific pathogen free (SPF) mice were intragastrically gavaged with 10^9 CFUs of the ExPEC cystitis isolate F11. At various time

points post gavage, feces were collected, and the number of viable ExPEC bacteria in the feces was enumerated. Throughout the course of the experiment, which lasted for 75 days after gavage, the median CFU per gram of feces was between 10^6 and 10^7, with the exception of a higher bacterial load at 1 day after gavage (Fig 5.1A). These data demonstrate that ExPEC can efficiently initiate and maintain colonization of the SPF mouse gut, as we have observed previously [23, 24]. Consistent with the observation that nonpathogenic *E. coli* mostly resides within the large bowel [25], the cecum and colon appear to carry the largest load of ExPEC bacteria, although there was a considerable amount of bacteria in the small intestine as well (Fig 5.1B).

It is of note that the SPF mice utilized in these experiments have not been treated with antibiotics and therefore carry an intact microbiota. This is in contrast to the streptomycin-treated mouse model, in which mice drink streptomycin water *ad libitum* prior to and during colonization [25]. The streptomycin-treated mouse model has been frequently used to study gut colonization by nonpathogenic *E. coli* strains, as it alters the microbiota in such a way as to clear a niche for a more consistent colonization phenotype. As streptomycin was not required for consistent ExPEC colonization, we wondered whether ExPEC is simply more adept at colonization than MG1655, a nonpathogenic *E. coli* strain commonly used with the streptomycin-treated mouse model. Indeed, when F11 and MG1655 were competed head-to-head in the mouse gut, the MG1655 bacteria were cleared from all of the mice by day 3, while F11 persisted (Fig 5.1C). When mice were gavaged with MG1655 alone, 80% of the mice were no longer colonized by day 14 (Fig 5.1D). In contrast, all of the mice that were gavaged with F11 alone remained colonized through day 14 (Fig 5.1D).

To determine whether ExPEC colonization is marked by inflammation, the mice were weighed at several time points after colonization as weight loss frequently accompanies colitis [26]. Mice gavaged with F11 consistently experienced transient

Figure 5.1. ExPEC colonizes the gut of SPF mice without causing serious pathology. Female Balb/c mice were gavaged at the age of 7-8 weeks with 10^9 CFUs of bacteria. (A) Mice were gavaged with F11, and fecal titers were determined at various time points. n=5 mice. (B) Mice were gavaged with F11, and the bacterial load for each organ of the gastrointestinal tract was determined. The mice were euthanized at 14 days post gavage and the stomach, small intestine (which was subsequently divided into three portions), the colon, and the cecum were removed from each mouse, homogenized, and serially diluted onto LB-clm plates. n=10 mice analyzed over two experiments. (C) Mice were gavaged with a 1:1 mixture of F11 and MG1655, and fecal titers were determined for both populations at various time points. ** $P \leq 0.01$ by Wilcoxon Signed-Rank Test. (D) Mice were gavaged with either F11, F11Δ*hlyA*, or MG1655, and fecal titers were determined at various time points. *** $P \leq 0.001$, **** $P \leq 0.0001$ by Mann-Whitney U Test. n=10 mice per group, analyzed over two experiments. (E) The mice in D were weighed at various time points. The values are normalized to the mass of each mouse prior to gavage. *** $P \leq 0.001$ when F11 is compared to F11Δ*hlyA*, or MG1655 by unpaired Student's t-test. (F) Mice were gavaged with either F11 or MG1655, and the colon was removed from each mouse 14 days post gavage. The colons were fixed and stained with H&E, and then imaged at 20x magnification. The scale bar indicates 100 μm. Images are representative of n=10 mice per group.


F11
F11ΔhlyA
MG1655
A
CFU/g
Days post inoculation
B
CFU/g
Stomach
Proximal
Middle
Distal
Colon
Cecum
Small Intestine
C
p=0.07
CFU/g
Days post inoculation
D
CFU/g
Days Post Inoculation
E
% of initial weight
Days Post Inoculation
F
F11
MG1655
100 µm

weight loss at 24 h in comparison to the mice colonized with MG1655 (Fig 5.1E). As the alpha-hemolysin toxin (HlyA) has been shown to induce inflammation in the gut in other mouse models [27, 28], we colonized mice with F11Δ*hlyA*. When mice were colonized with F11Δ*hlyA*, the bacterial load was lower at 1 and 3 days post inoculation, with a 10-fold reduction in bacterial levels at day 1 (Fig 5.1D). Additionally, transient weight loss did not occur in mice colonized with F11Δ*hlyA* (Fig 5.1E), suggesting that HlyA is required for both full bacterial levels and short-term inflammation during the initiation of colonization. The idea that the effect of this inflammation is short-lived is supported by the observation that at day 14, the colons of mice colonized with F11 retain their normal crypt architecture, and do not contain infiltrating immune cells (Fig 5.1F). Altogether, these data suggest that F11 induces transient inflammation that manifests as temporary weight loss, a condition that is not sustained after 24 h of colonization.

Various ExPEC toxins are not important for colonization of the mouse gut

In light of the hypothesis of coincidental evolution, it is intriguing that a canonical ExPEC virulence factor, such as HlyA, is not required for sustained colonization of the mouse gut (Fig 5.1D). To test whether other ExPEC toxins could be involved in gut colonization, we examined cytotoxic necrotizing factor type 1 (Cnf1), uropathogenic specific protein (Usp), and colibactin.

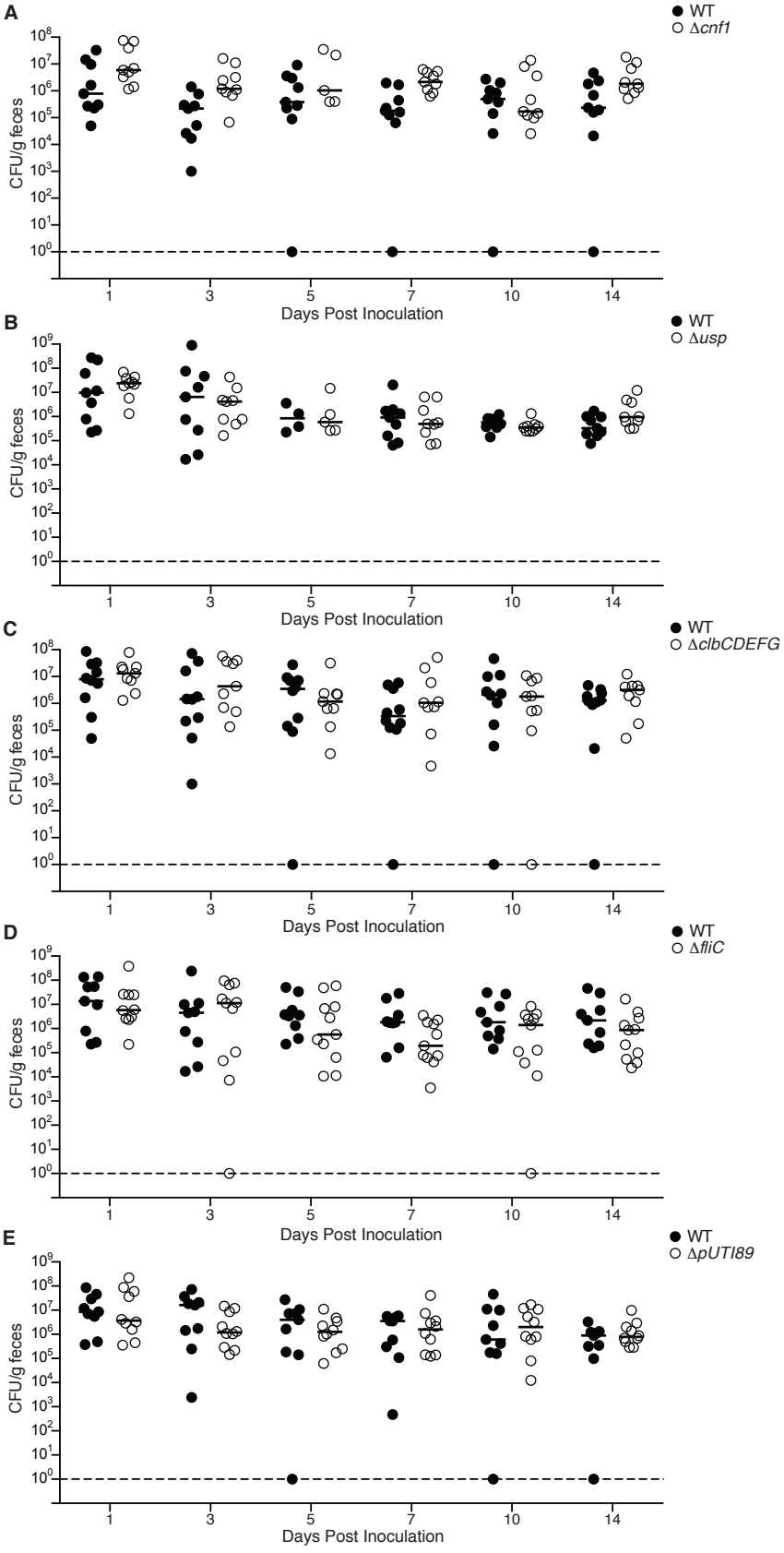
Cnf1 intoxicates host cells via deamidation of a specific glutamine amino acid of various Rho family GTPases such as RhoA, Rac, and Cdc42, which leads to constitutive activation of these proteins [29]. The activation of Rho GTPases by Cnf1 leads to rearrangements of the host cytoskeleton and multinucleation [30, 31]. The actions of Cnf1 induce an inflammatory response in both UTIs [32] and prostatitis [33]. Although Cnf1 mediates ExPEC virulence, the contribution of Cnf1 to bacterial fitness remains

unclear, as it is important for host colonization in some cases [34] but not others [32, 33, 35]. Given the association of Cnf1 with ExPEC in epidemiological studies [31], and its potential as an ExPEC virulence and fitness factor, we tested whether it plays a role in the gut. After feeding mice either 10^9 CFUs of WT F11 or F11Δ*cnf1*, bacterial colonization levels were tracked by homogenizing and plating feces at various time points. Median bacterial levels ranged between 10^5 and 10^6 CFU per gram of feces (Fig 5.2A). Overall, there was no change in colonization levels of the F11Δ*cnf1* strain in comparison to WT (Fig 5.2A), suggesting that Cnf1 does not contribute to the ability of F11 to colonize the mouse gut.

Like many of the other ExPEC virulence factors, Usp has been linked to ExPEC epidemiologically, meaning that the *usp* gene is more often found in ExPEC isolates than in commensal isolates of *E. coli* [36, 37]. Usp was also associated with *E. coli* fecal isolates that were capable of persisting in the guts of infants and that carried other ExPEC markers [38]. The physiological function of Usp is not clear, although it has been proposed to be genotoxic, as it is a nuclease that can intoxicate host cells *in vitro* [39, 40]. The sequence of Usp indicates that it could also act as a colicin, a toxin that will harm unrelated bacteria, which is strengthened by the observation that several genes downstream of *usp* have sequence similar to colicin immunity proteins [41]. Given that interbacterial competition is expected to occur in the gut, we tested whether Usp is important for ExPEC gut colonization. The F11Δ*usp* mutant did not exhibit a defect during gut colonization, indicating that it is not required for this niche (Fig 5.2B).

Another toxin that has been linked to ExPEC pathogenesis is colibactin, which is produced by the polyketide synthase (*pks*) genomic island. The *pks* island is not found in intestinal pathogenic *E. coli* strains, but is enriched in extraintestinal isolates in comparison to fecal commensal isolates [42]. Colibactin induces DNA damage and cell cycle arrest in host cells [42, 43], and the presence of colibactin-producing bacteria has

Figure 5.2. Several ExPEC virulence factors are not required for gut colonization. Mice were gavaged with 10^9 CFUs of WT or mutant bacteria, including (A) Δ*cnf1*, (B) Δ*usp*, (C) Δ*clbCDEFG*, (D) Δ*fliC*, and (E) Δ*pUTI89*. Feces were collected at the indicated time points and titered to enumerate the bacterial load in the feces. n = 9-11 mice per group. No statistical significance found by comparing WT and mutant titers at every time point via Mann-Whitney U test.


A
WT
Δcnf1
CFU/g feces
Days Post Inoculation
B
WT
Δusp
CFU/g feces
Days Post Inoculation
C
WT
ΔclbCDEFG
CFU/g feces
Days Post Inoculation
D
WT
ΔfliC
CFU/g feces
Days Post Inoculation
E
WT
ΔpUTI89
CFU/g feces
Days Post Inoculation

been linked to colorectal cancer [44]. In extraintestinal infections, colibactin-deficient bacterial strains are not as virulent as their WT counterparts. Mutation of the *pks* island reduces ExPEC translocation from the gut in neonatal rats [45] and reduces lymphopenia in septic mice [46]. Whether the *pks* island also plays a role in bacterial fitness during gut colonization is not clear. Whereas one study observed decreased gut colonization by colibactin mutants [45], another found no difference between the mutant and WT bacteria [47]. To test whether colibactin is important in our model of gut colonization, we colonized mice with either WT F11 or F11Δ*clbCDEFG*—removing the entirety of one of the *pks* island operons [48]—and tracked fecal titers over time. There was no difference in the colonization ability of the WT and mutant strains (Fig 5.2C), suggesting that colibactin does not contribute to bacterial fitness in this model of gut colonization.

Flagella are not required for ExPEC gut colonization

Flagella create motility through a membrane-localized motor that drives an extracellular filament composed mostly of FliC [49]. Several studies have provided evidence that flagella play an important role during ExPEC infection of the mouse urinary tract [50-53]. The expression of *fliC*, which encodes for the major flagellar protein, is spatiotemporally regulated during UTI. It is expressed early during infection in the bladder, but as the infection progresses, the expression in the bladder decreases while the expression in the kidneys increases, suggesting that flagellar expression is required for ascension of the urinary tract [53]. Although flagella are not important for gut colonization of the streptomycin-treated mouse by nonpathogenic *E. coli* [54], it was not known whether flagella are important for gut colonization by ExPEC in the face of an intact microbiota. The F11Δ*fliC* KO mutant was able to colonize the mouse gut as well as the WT strain (Fig 5.2D), suggesting that flagella do not play an important role for

ExPEC in the gut.

The pUTI89 plasmid is not important for gut colonization

The pUTI89 plasmid was first identified when the cystitis isolate UTI89 was sequenced [55]. It is roughly 114 kb in length, and encodes conjugation machinery, some virulence genes, and many hypothetical genes. A derivative of UTI89 that was cured of the plasmid had reduced fitness at early time points of urinary tract infection, in addition to a defect in invasion of host cells and formation of intracellular bacterial communities [56]. Other ExPEC strains have since been found to carry identical or closely-related plasmids [57], including F11, in which disruptions of several plasmid genes can lead to *in vitro* growth defects (see Chapter 3). To determine the role that pUTI89 plays during gut colonization, an F11 derivative cured of the plasmid was tested in our mouse model. The cured strain exhibited no defect when compared to the control WT strain (Fig 5.2E), suggesting that pUTI89 is not important for fitness within the gut.

Type 1 Pili contribute to fitness in the gut, whereas the P pili adhesin is dispensable for gut colonization

ExPEC encodes many types of pili, which are hair-like fibers that extend from the cell and terminate in an adhesin protein that interacts with its surrounding environment. For type 1 pili (T1P), this adhesin is FimH, while PapG is the adhesin for P pili. T1P have been shown to be important for many steps in the urinary tract infection process, including adherence to bladder epithelial cells, subsequent invasion into host cells, and formation of intracellular bacterial communities [58]. While T1P play a role in the bladder, P pili mediate binding to the kidney epithelium and are also important for ExPEC-mediated UTI [58].

To determine whether these adhesins are important for successful gut colonization, each adhesin was disrupted and the resulting mutants were tested in

competitive colonization experiments. The WT strain and either F11Δ*fimH* or F11Δ*papG* were mixed in equal amounts and fed to mice, and bacterial titers were enumerated at various time points post inoculation. A control experiment was also performed to determine the relative fitness effects of the antibiotic resistance cassettes. In this control experiment, the kan^R WT strain exhibited somewhat of a defect (about 3-fold with a median competitive index of -0.49) in comparison to the clm^R WT strain at 14 days post inoculation (Fig 5.3A). Furthermore, 2 of the 13 mice cleared the WT-kan strain by day 14, as indicated by the green data points (Fig 5.3A). The WT vs Δ*papG* experiment demonstrated a transient defect at day 5 with a median competitive index of -0.83 (6.8-fold mutant defect), but no defect was observed at later time points (Fig 5.3B). By day 14, one mouse had cleared the kan^R Δ*papG* mutant (green dot) while one mouse had cleared the clm^R WT competitor (red dot). In contrast, the kan^R Δ*fimH* mutant had significantly reduced fitness beginning at day 7 and continuing throughout the rest of the experiment (Fig 5.3C), with the greatest defect seen at day 10 (median competitive index of -2.55, indicating a ~350-fold mutant defect). The mice started clearing the Δ*fimH* mutant as early as day 3, with 6 of 13 mice clearing the mutant by day 10 (green dots). These data indicate that T1P may play a role in gut colonization, whereas P pili do not.

These observations are supported by noncompetitive experiments in which WT and mutant strains were inoculated individually into separate groups of mice. The Δ*papG* mutant strain colonized as well as WT (Fig 5.3D), whereas the Δ*fimH* mutant was cleared in 3 of the 10 mice (Fig 5.3E), a difference that did not reach statistical significance.

Disruption of fimH or papG alters bacterial motility

As alteration of T1P and P pili expression is known to lead to modifications in motility [59-61], we tested the ability of the adhesin mutant strains to swim in motility

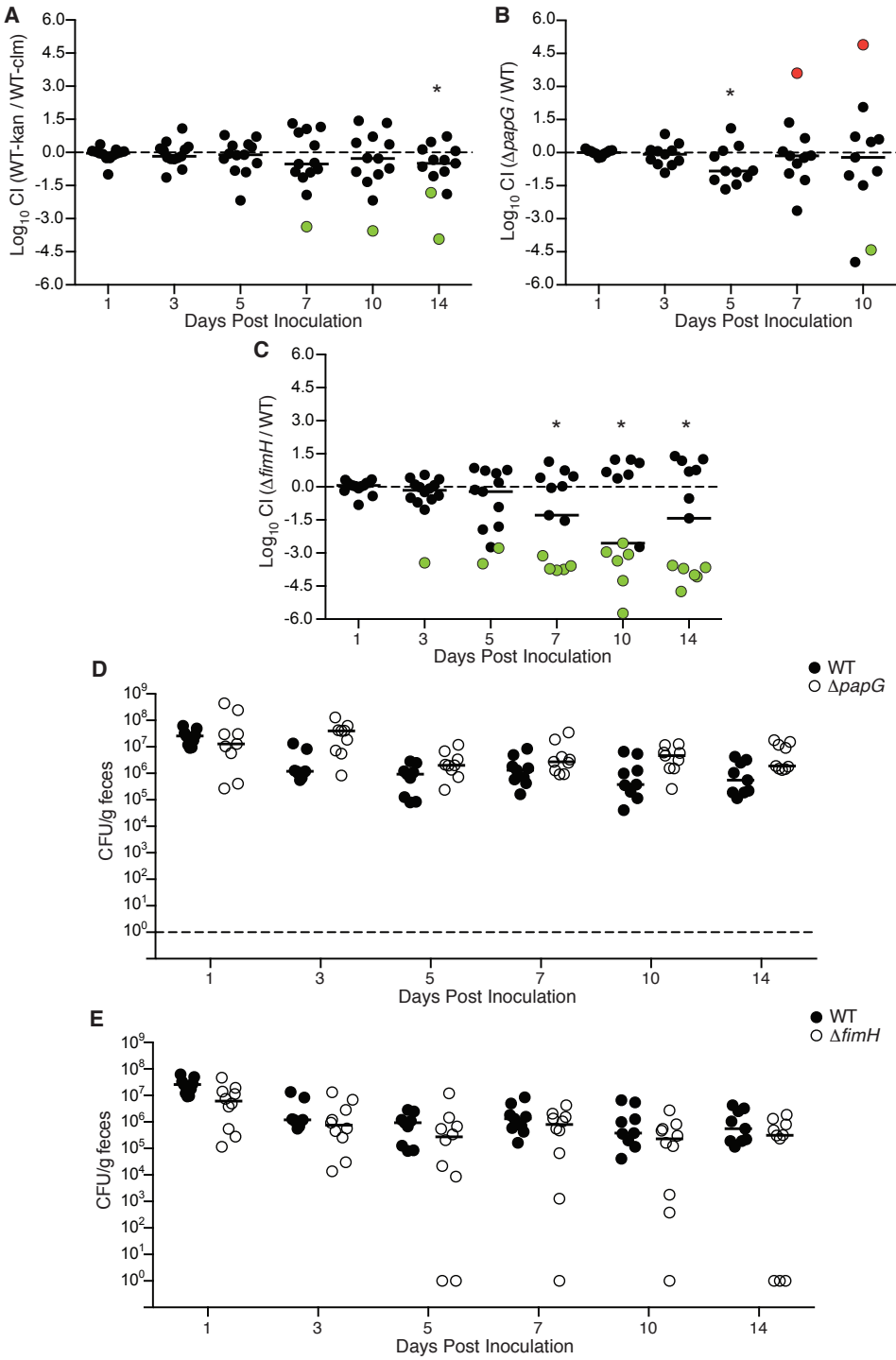


Figure 5.3. The Δ*fimH* mutant, but not a Δ*papG* mutant, has a defect in the gut. Mice were gavaged with 10^9 CFUs of a mixture of (A) WT::kan and WT::clm, (B) WT and Δ*papG*, (C) or WT and Δ*fimH*. n = 11-13 mice per competition. * $P \leq 0.05$ by one sample t test. Mice were gavaged with WT bacteria and either (D) Δ*papG* or (E) Δ*fimH*. No statistical significance found by comparing WT and mutant titers at every time point via Mann-Whitney U test.

agar. The Δ*fimH* mutant had increased motility when compared to the WT strain, whereas the Δ*papG* strain exhibited decreased motility (Fig 5.4A).

Aberrant overexpression of flagella can lead to a defect in gut colonization [5], and mutants expressing lower levels of flagella can spontaneously arise over the course of gut colonization experiments [62, 63]. To test whether the defect in the Δ*fimH* mutant strain was due to increased flagella, an F11Δ*fimH*Δ*fliC* double knockout strain was created. This strain exhibited no motility, similar to the single Δ*fliC* KO (Fig 5.4B). During competitive gut colonization, the clm^R Δ*fimH*Δ*fliC* mutant strain exhibited a defect in comparison to the kan^R WT strain (Fig 5.4C). By day 14, 5 of the 13 mice had cleared the mutant strain, and the median competitive index was -1.33. However, this defect did not appear to be as strong as was observed with the Δ*fimH* mutant, as the reduction in the competitive index was smaller and not statistically significant. Clearance of the Δ*fimH*Δ*fliC* also began later, at day 7 (Fig 5.4C), in comparison to the Δ*fimH* strain that began to be cleared by the mice at day 3 (Fig 5.3C). Considered together, these data suggest that the defect of the Δ*fimH* strain is due to both increased motility and the lack of T1P.

The fim switch is mostly in the OFF state in feces-localized ExPEC

There are many levels of regulation that modulate the expression of T1P. One aspect of T1P regulation is that the *fim* switch—a short span of DNA containing the promoter for the main *fim* operon—can be flipped back and forth by various recombinases [58]. In this manner, environmental cues can influence whether the *fim* switch is in the ON state and primed for transcription, or whether it is in the OFF state, in which transcription is impossible. To quantify the percentage of the ExPEC gut population in which the switch is ON or OFF, DNA was isolated from the feces of mice

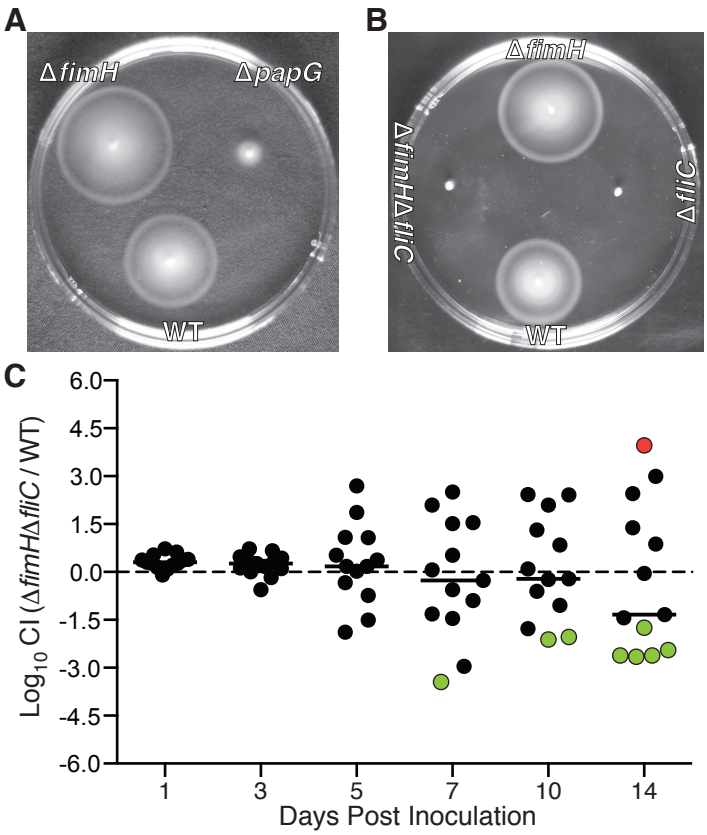


Figure 5.4. The increased motility of the Δ*fimH* mutant contributes to its gut colonization defect. Bacteria were inoculated into motility agar to observe their swimming ability. WT bacteria as well as (A) Δ*fimH* and Δ*papG*, or (B) Δ*fimH*, Δ*fliC*, and Δ*fimH* Δ*fliC*, were tested. Images are representative of 3 experiments. (C) Mice were gavaged with a mixture of 10^9 CFUs of WT and Δ*fimH* Δ*fliC* bacteria. n = 13 mice. No statistical significance via one sample t test.

that were colonized with WT F11 (Fig 5.5). A PCR reaction that spans the *fim* switch was performed, followed by a restriction enzyme digest with HinfI, which cuts in one location within the *fim* switch and results in different sized products depending on the state of the switch. As expected, growing F11 in pure culture led to different switch states depending on the growth conditions (Fig 5.5). The majority of the population carried an ON switch when the bacteria were grown statically, whereas shaking incubation led to switches that were mostly OFF (Fig 5.5). Prior to gavage, no *fimS* was detected in the feces. At various points post gavage, this assay revealed that the majority of the population (~70%, on average) carried an OFF switch. The OFF state of the *fim* switch observed during ExPEC colonization of mice was recapitulated in a human fecal sample from a patient colonized with at least two ExPEC strains. These data indicate that only a portion of the ExPEC fecal population is capable of expressing T1P.

T1P facilitate the resurgence of ExPEC after antibiotic treatment

One roadblock for treatment of ExPEC-mediated UTI is the issue of recurrence. Frequently, women who suffer a UTI will contract another UTI relatively shortly after the initial infection is resolved [64]. It is thought that the gut may act as a reservoir that reseeds the urinary tract, therefore mediating the recurrent infection. An unsettling observation is that the ExPEC gut population can either rebound shortly after the termination of antibiotic treatment, or is never cleared by the antibiotic treatment in the first place [15]. How the ExPEC gut population is able to tolerate antibiotic treatment has not been studied, but a similar phenotype can be observed during bladder colonization. ExPEC is capable of utilizing T1P to invade into bladder epithelial cells, and within this intracellular niche, ExPEC is protected from membrane impermeable antibiotics [65].

To test whether antibiotic tolerance in the gut could also be mediated by T1P, we tested the F11 WT strain and the Δ*fimH* mutant strain in a mouse model of antibiotic

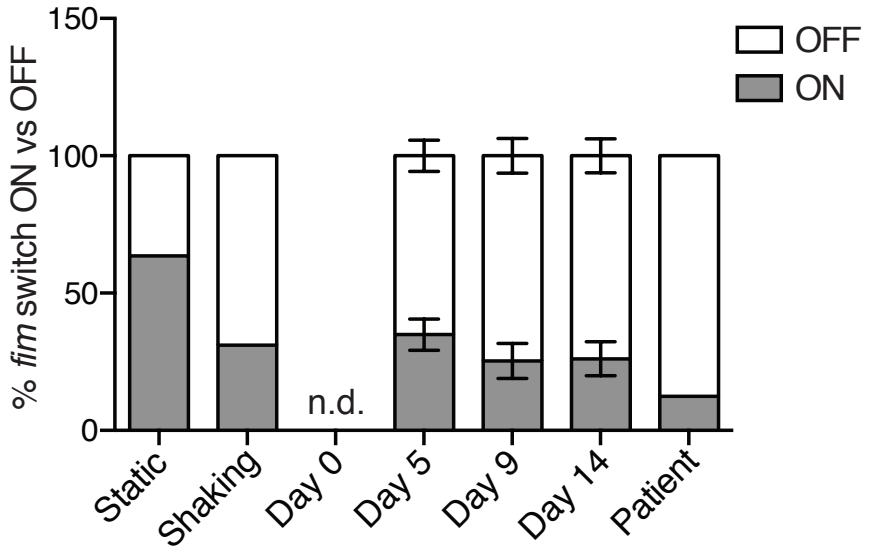


Figure 5.5. The *fim* switch is mostly OFF within the ExPEC fecal population. Mice were gavaged with WT F11 and fecal samples were acquired at various time points post gavage. As a control, fecal samples were also collected prior to gavage. DNA was isolated from each sample, and the *fim* switch status in the population was quantified using a PCR that spanned the switch, followed by HinfI digestion and measurement of ON and OFF bands on an agarose gel. For controls, the bacteria were grown in LB for either 72 h statically (ON), or shaking until exponential phase (OFF). Feces from a patient carrying at least two distinct ExPEC populations in her gut was also analyzed.

treatment. Mice were gavaged with WT bacteria, the Δ*fimH* mutant, or PBS. At day 7 post gavage, no ExPEC were detected in the PBS-gavaged mice as expected (Fig 5.6A), whereas similar amounts of WT and mutant F11 bacteria were observed (Figs 5.6B-5.6C). The membrane-impermeable antibiotic gentamicin was added to the drinking water at day 7, and after 3 days of treatment, no ExPEC bacteria were detected in any of the mice (Figs 5.6A-5.6C). Antibiotic treatment was halted on day 11, and 24 h later, the ExPEC population rebounded in all of the mice gavaged with WT bacteria (Fig 5.6B). In contrast, the Δ*fimH* mutant strain reemerged in only 1 of the 4 mice, and in that case, the bacteria became detectable 48 h after the removal of gentamicin (Fig 5.6C). Importantly, no ExPEC bacteria were detected in the PBS control mice at any time (Fig 5.6A). Furthermore, the mice drank equal amounts of water prior to and during gentamicin treatment (Fig 5.6D). These data indicate that T1P play a role in the ability of ExPEC to reestablish a population within the gut post antibiotic treatment.

Discussion

As the gastrointestinal tract is the major host niche for ExPEC, the gut environment could be the most prominent driver of ExPEC evolution. With this in mind, it has been hypothesized that the extraintestinal success of ExPEC is a by-product of its ability to colonize the gut [21]. This hypothesis would be supported by evidence that extraintestinal virulence or fitness factors also play a role in gut colonization. Therefore, we set out to test the relevance of several canonical ExPEC virulence factors in a mouse model of gut colonization. We found that several factors played no role during colonization, including Cnf1, Usp, colibactin, flagella, and the plasmid pUTI89 (Fig 5.2). Mutants lacking either *papG* or *hlyA* exhibited only transient colonization defects that did not affect overall gut survival (Figs 5.1D & 5.3B). In contrast, T1P were important for gut colonization and persistence (Figs 5.3C & 5.3E). These data indicate that some

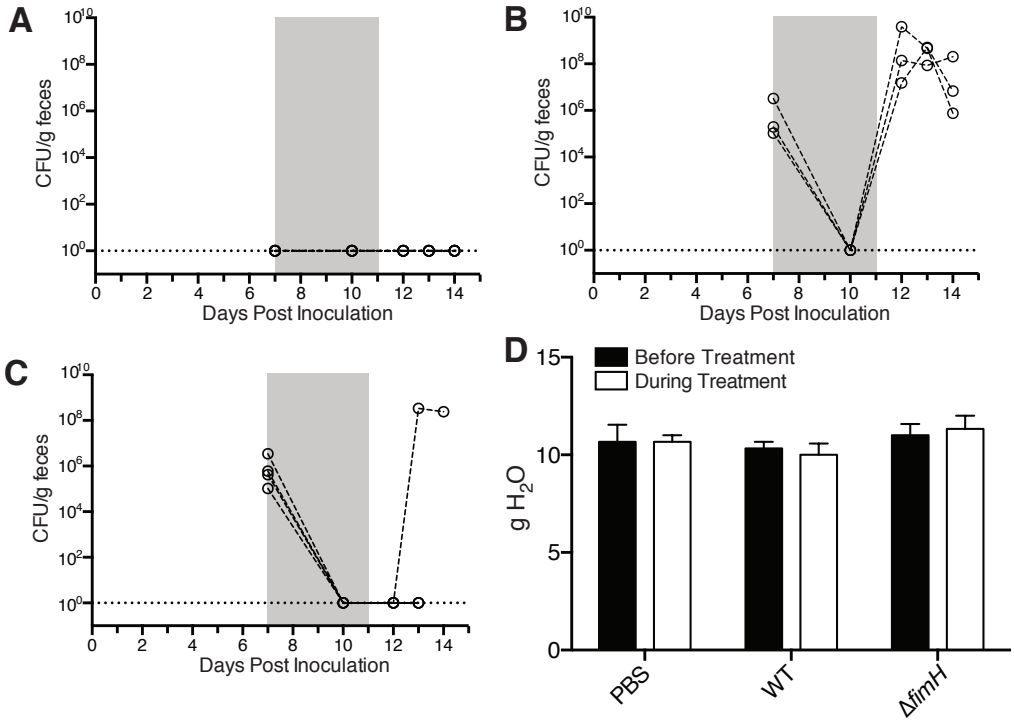


Figure 5.6. T1P contribute to the ability of ExPEC to reemerge after oral antibiotic treatment. Mice were gavaged with (A) PBS, or 10^9 of either (B) WT F11 or (C) F11Δ*fimH* bacteria. At several time points post gavage, feces were collected, homogenized, and plated onto MacConkey agar in order to enumerate the ExPEC population. Collection was done prior to, during, and after gentamicin treatment, which began on day 7 and ended on day 11, and which is indicated by the gray box. PBS, n = 5 mice. WT, n = 3 mice. Δ*fimH*, n = 4 mice. (D) The amount of water consumed per day by each group of mice was measured prior to and during gentamicin treatment to ensure equal water intake.

ExPEC virulence factors are important for gut colonization, while others are not.

Therefore, it is reasonable to apply the principle of coincidental evolution to individual genes, but there are likely many forces from various niches that have shaped ExPEC genetics over time. For example, one study found a correlation between extraintestinal virulence and the ability to resist killing by amoebae, highlighting that diverse environmental situations could have influenced extraintestinal pathogenesis [66].

A previous study examined ExPEC coincidental evolution by testing the importance of the 7 major pathogenicity islands of the cystitis isolate 536 during extraintestinal infection and gut colonization [5]. It was found that a mutant that lacked the 7 pathogenicity islands had reduced virulence in a mouse model of sepsis as well as in gut colonization. However, the gut colonization defect was attributed to aberrantly increased flagellar expression, a condition that is detrimental to *E. coli* fitness in the gut [62, 63]. This is similar to our observation that the gut colonization defect of the Δ*fimH* mutant could be partially attributed to its increased motility (Fig 5.4A). Importantly, the Δ*fimH* Δ*fliC* double knockout strain continued to have a gut colonization defect (Fig 5.4C), although it was more modest than that exhibited by the Δ*fimH* mutant (Fig 5.3C).

In addition to being important for persistence in the gut, T1P also play a role in the reestablishment of the ExPEC population after an antibiotic treatment (Fig 5.6C). This has implications in the realm of recurrent UTIs (rUTIs), which can be a major challenge in UTI treatments [67]. These rUTIs are often thought to be due to a reservoir of ExPEC within the gut that repeatedly supplies bacteria to the urinary tract. This hypothesis assumes that the gut ExPEC reservoir is either not efficiently cleared by the antibiotic used to treat the UTI, or that gut colonization is easily occurring multiple times. The former possibility is supported by a study indicating that ExPEC can be detected within the gut between two UTI episodes [15]. In that study, two patterns were

observed, one in which the ExPEC population was unaltered by the antibiotic treatment, and another in which the population was cleared but later reemerged. In our mouse model, we observed that the bacteria became undetectable after 3 days of oral gentamicin treatment, but quickly rebounded (Fig 5.6). Why T1P contributed to this phenotype remains unclear, but it could be related to the ability of the bacteria to access a niche that would otherwise be unavailable. For example, in the urinary tract, T1P allow the bacteria to invade into the bladder epithelial cells, and within this intracellular compartment, the bacteria are less affected by antibiotics [65]. Whether invasion of host cells also occurs in the gut is not currently clear, but such a niche would allow a subset of the bacteria to survive the gentamicin treatment and repopulate the gut.

Another interesting possibility that could explain the role of T1P in the gut is a scenario in which T1P and secretory IgA (sIgA) interact to create an ExPEC biofilm that is resistant to antibiotics. This idea is supported by the fact that T1P can bind to the heavily glycosylated sIgA proteins [68], and that this interaction can enhance biofilm formation *in vitro* [69]. Interestingly, there is a correlation between the production of sIgA in humans and the presence of *E. coli* that is capable of T1P-mediated adherence [70]. *E. coli* isolated from individuals that are deficient in sIgA has a reduced capacity for T1P-mediated adherence in comparison to *E. coli* from healthy volunteers. This difference can be attributed to many factors, including fewer bacteria carrying the *fim* genes, a reduced ability to switch to the ON phase, and possibly ineffective *fimH* alleles [70];Nowrouzian, 2007 #60}. These data suggest that there is an association of T1P+ bacteria with sIgA-producing individuals, while T1P- bacteria are more often found with sIgA-deficient individuals. The co-occurrence of T1P and sIgA could also be a clue for the function of T1P in the absence of antibiotics.

A striking feature of our mouse model is that ExPEC is able to colonize SPF mice that carry an intact microbiota, which has also been observed by others [71]. However, *E.*

coli gut colonization is traditionally studied using mice that have been treated with streptomycin in order to provide an open niche [72]. Disturbance of the microbiota has been commonly utilized as *E. coli* is unable to experimentally colonize a gut with an intact microbiota. For example, when volunteers ingested a marked derivative of *E. coli* isolated from their own guts, the strains were unable to persist [73]. This failure to colonize can be attributed to colonization resistance, which is due—in part—to the microbiota reducing niche availability and restricting colonization by incoming bacteria [74]. It therefore seems likely that as microbiota composition varies from one facility to the another [75], some mice are amenable to ExPEC colonization due to the specific makeup of their microbiota. Indeed, the mice in our facility do not appear to be colonized with *E. coli* prior to beginning our experiments, which may indicate a niche is available for ExPEC to fill.

The results of this work indicate that T1P are important for gut colonization, and for repopulating the gut after antibiotic treatment. Furthermore, we found that coincidental evolution was a feasible explanation for the existence of only one of several ExPEC virulence factors, indicating that piecing together the evolutionary history of these bacteria is a complicated task. Future work to label genes as gut-specific, extraintestinal-specific, environment-related, or multi-niche—especially in a high-throughput manner—will better paint a picture of the pressures experienced by ExPEC.

Methods

Bacterial strain creation

The cystitis isolate F11 and the K-12 strain MG1655 were genetically modified using lambda-red recombination that was facilitated by the pKM208 plasmid [76]. Most of the constructs used for recombination were created by PCR using either pKD4 or pKD3 as a template to make a kanamycin or chloramphenicol resistance cassette,

respectively, flanked with ~40 bps of homology to the insertion site. In some cases, longer homology regions were required, and three-part PCR was performed. This was done by PCR of the upstream region, the resistance cassette, and the downstream region, all with some homology to the adjacent piece. The three pieces were then stitched together in a single reaction.

The pCP20 plasmid was used to remove the resistance cassette when needed [77]. To cure F11 of the pUTI89 plasmid, the *ccdAB* toxin-antitoxin system was replaced with a *tetA-sacB* construct and spontaneous loss of the plasmid was selected for on LB plates containing fusaric acid and sucrose, as explained before [78].

Media

Prior to lambda-red recombination, bacteria were grown in LB broth. All growth in petri dishes was done using LB agar supplemented with chloramphenicol (20 μg/mL) or kanamycin (50 μg/mL). Prior to experiments, bacteria were grown in a modified M9 medium, which contains the following in water: $MgSO_4 \bullet 7H_2O$ (1 mM), $CaCl_2 \bullet 2H_2O$ (0.1 mM), D-(+)-Glucose (0.1%), Nicotinic Acid (0.00125%), Thiamine HCl (0.00165%), Casamino Acids (0.2%), Na_2HPO_4 (6g/L), KH_2PO_4 (3 g/L), NH_4Cl (1 g/L), and NaCl (0.5 g/L).

Mouse gut colonization

Mice were handled in accordance with protocols approved by the Institutional Animal Care and Use Committee at the University of Utah (Protocol number 15-12015), following US federal guidelines indicated by the Office of Laboratory Animal Welfare (OLAW) and described in the Guide for the Care and Use of Laboratory Animals, 8th Edition.

To prepare bacteria for inoculation into mice, the bacteria were grown in 20 mL of modified M9 media in a 250 mL flask for 24 h at 37° C, without shaking. A total of 12

mL of culture (6 mL of each culture for competitive experiments) was centrifuged at 8,000 r.c.f. for 10 min, and the pellet was washed and resuspended in 0.5 mL of PBS. To colonize the mouse gastrointestinal tract, 7-8-week-old female Balb/c mice were intragastrically gavaged with 10^9 CFUs of bacteria in a volume of 50 µl. At various time points post inoculation, mice were placed individually into unused takeout boxes for weighing and feces collection. Feces were homogenized in 1 mL of 0.7% NaCl, briefly centrifuged to pellet insoluble debris, and the supernatants were serially diluted and spread onto LB plates containing either chloramphenicol (20 µg/mL) or kanamycin (50 µg/mL) to select for the relevant strains. Prior to gavage, fecal samples were analyzed to make sure there were no endogenous bacteria resistant to chloramphenicol or kanamycin. The mice were housed 3-5 per cage, and were allowed to eat (Teklad Global Soy Protein-Free Extruded, irradiated, 2920x) and drink antibiotic-free water *ad libitum*.

To acquire a count of the bacterial burden in various organs of the gastrointestinal tract, the mice were anaesthetized via isoflurane and euthanized by cervical dislocation. The cecum, colon, small intestine, and stomach were removed. The small intestine was divided into thirds, with the portion closest to the stomach labeled “proximal”, and the portion closest to the cecum labeled “distal”. A part of each organ was then weighed and placed into a Safe-Lock tube (Eppendorf) with three 3.2 mm stainless steel beads, and homogenized using the Bullet Blender (Next Advance). The homogenate was then serially diluted and plated onto LB plates containing chloramphenicol (20 µg/mL) to quantify the levels of F11::clm bacteria. For histology, a portion of the colon was removed, placed into a cassette, and fixed in 10% neutral buffered formalin and submitted to the University of Utah Research Histology group for processing, which included hematoxylin and eosin staining.

Gentamicin treatment assay

Mice were gavaged as described above with either 10^9 CFUs of WT F11 or F11Δ*fimH*, or with PBS. Seven days post inoculation, the feces were titered onto MacConkey agar to quantify the bacterial load. On day 7, gentamicin (Sigma #G1264) was added to the water at 1 g/L, and was removed on day 11 after confirming clearance of ExPEC on day 10. More fecal samples were analyzed at later time points to observe whether the bacterial population grew out again.

Quantification of the fim switch status

Mice were gavaged as described above with 10^9 CFUs of WT F11. Fecal samples were collected prior to gavage and at various time points post gavage, and DNA was isolated using the ZR Fecal DNA MiniPrep by Zymo Research. The patient fecal sample was processed in parallel with the same kit. As a control for a population with the majority of the switch ON, WT F11 was grown in LB statically for 72 h, with 1:100 subculturing done at 24 and 48 h. As an OFF control, WT F11 was grown shaking in LB to exponential phase. For both controls, DNA was isolated using the Promega Wizard Genomic DNA Purification Kit. For each sample, a PCR reaction was run using primers that spanned the *fim* switch (F: 5'-TACCGCCAGTAATGCTGCTC-3', R: 5'-GTCCCACCATTAACCGTCGT-3'). The extension portion of the reaction lasted for 25 cycles, with a T_m of 56° C. Each reaction was purified, and then digested with HinfI for 1.5 h at 37° C, after which each reaction was run on a 2% agarose gel. The intensities of the ~560 bp (ON) and ~430 bp (OFF) bands were analyzed in ImageJ.

Motility assay

To test the swimming ability of particular strains, motility agar plates were made by pouring 25 mL of LB soft agar (0.1% agar) into a petri dish. A small volume of bacteria (1.5 µl) from an overnight culture was dispensed just below the surface of the plate. The

plate was incubated at 37° C for 8-10 h prior to imaging with the Stratagene Eagle Eye II Imaging System.

Statistical analysis

All analyses were done in GraphPad Prism. The specific statistical test used in each case is indicated in the figure legend.

References

1. Blount ZD. The unexhausted potential of E. coli. Elife. 2015;4. doi: 10.7554/eLife.05826. PubMed PMID: 25807083; PubMed Central PMCID: PMCPMC4373459.

2. Bezirtzoglou E. The intestinal microflora during the first weeks of life. Anaerobe. 1997;3(2-3):173-7. doi: 10.1006/anae.1997.0102. PubMed PMID: 16887585.

3. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486(7402):207-14. doi: 10.1038/nature11234. PubMed PMID: 22699609; PubMed Central PMCID: PMCPMC3564958.

4. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal Escherichia coli. Nat Rev Microbiol. 2010;8(3):207-17. doi: 10.1038/nrmicro2298. PubMed PMID: 20157339.

5. Diard M, Garry L, Selva M, Mosser T, Denamur E, Matic I. Pathogenicity-associated islands in extraintestinal pathogenic Escherichia coli are fitness elements involved in intestinal colonization. J Bacteriol. 2010;192(19):4885-93. doi: 10.1128/JB.00804-10. PubMed PMID: 20656906; PubMed Central PMCID: PMCPMC2944530.

6. Karpman D, Stahl AL. Enterohemorrhagic Escherichia coli Pathogenesis and the Host Response. Microbiol Spectr. 2014;2(5). doi: 10.1128/microbiolspec.EHEC-0009-2013. PubMed PMID: 26104359.

7. Rolhion N, Darfeuille-Michaud A. Adherent-invasive Escherichia coli in inflammatory bowel disease. Inflamm Bowel Dis. 2007;13(10):1277-83. doi: 10.1002/ibd.20176. PubMed PMID: 17476674.

8. Schultz M, Strauch UG, Linde HJ, Watzl S, Obermeier F, Gottl C, et al. Preventive effects of Escherichia coli strain Nissle 1917 on acute and chronic intestinal inflammation in two different murine models of colitis. Clin Diagn Lab Immunol. 2004;11(2):372-8. PubMed PMID: 15013990; PubMed Central PMCID: PMCPMC371200.

9. Deriu E, Liu JZ, Pezeshki M, Edwards RA, Ochoa RJ, Contreras H, et al. Probiotic bacteria reduce salmonella typhimurium intestinal colonization by competing for iron. Cell Host Microbe. 2013;14(1):26-37. doi: 10.1016/j.chom.2013.06.007.

PubMed PMID: 23870311; PubMed Central PMCID: PMCPMC3752295.

10. Vila J, Saez-Lopez E, Johnson JR, Romling U, Dobrindt U, Canton R, et al. Escherichia coli: an old friend with new tidings. FEMS Microbiol Rev. 2016. doi: 10.1093/femsre/fuw005. PubMed PMID: 26961094.

11. Barber AE, Norton JP, Spivak AM, Mulvey MA. Urinary tract infections: current and emerging management strategies. Clin Infect Dis. 2013;57(5):719-24. doi: 10.1093/cid/cit284. PubMed PMID: 23645845; PubMed Central PMCID: PMCPMC3739462.

12. Smati M, Clermont O, Le Gal F, Schichmanoff O, Jaureguy F, Eddi A, et al. Real-time PCR for quantitative analysis of human commensal Escherichia coli populations reveals a high frequency of subdominant phylogroups. Appl Environ Microbiol. 2013;79(16):5005-12. doi: 10.1128/AEM.01423-13. PubMed PMID: 23770894; PubMed Central PMCID: PMCPMC3754702.

13. Nowrouzian FL, Wold AE, Adlerberth I. Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. J Infect Dis. 2005;191(7):1078-83. doi: 10.1086/427996. PubMed PMID: 15747243.

14. Nowrouzian FL, Adlerberth I, Wold AE. Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes Infect. 2006;8(3):834-40. doi: 10.1016/j.micinf.2005.10.011. PubMed PMID: 16483819.

15. Russo TA, Stapleton A, Wenderoth S, Hooton TM, Stamm WE. Chromosomal restriction fragment length polymorphism analysis of Escherichia coli strains causing recurrent urinary tract infections in young women. The Journal of infectious diseases. 1995;172(2):440-5. Epub 1995/08/01. PubMed PMID: 7622887.

16. Moreno E, Andreu A, Pigrau C, Kuskowski MA, Johnson JR, Prats G. Relationship between Escherichia coli strains causing acute cystitis in women and the fecal E. coli population of the host. J Clin Microbiol. 2008;46(8):2529-34. doi: 10.1128/JCM.00813-08. PubMed PMID: 18495863; PubMed Central PMCID: PMCPMC2519474.

17. Chen SL, Wu M, Henderson JP, Hooton TM, Hibbing ME, Hultgren SJ, et al. Genomic diversity and fitness of E. coli strains recovered from the intestinal and urinary tracts of women with recurrent urinary tract infection. Sci Transl Med. 2013;5(184):184ra60. doi: 10.1126/scitranslmed.3005497. PubMed PMID: 23658245; PubMed Central PMCID: PMCPMC3695744.

18. Foxman B, Manning SD, Tallman P, Bauer R, Zhang L, Koopman JS, et al. Uropathogenic Escherichia coli are more likely than commensal E. coli to be shared between heterosexual sex partners. Am J Epidemiol. 2002;156(12):1133-40. PubMed PMID: 12480658.

19. Yamamoto S, Tsukamoto T, Terai A, Kurazono H, Takeda Y, Yoshida O. Genetic evidence supporting the fecal-perineal-urethral hypothesis in cystitis caused by Escherichia coli. J Urol. 1997;157(3):1127-9. PubMed PMID: 9072556.

20. Levin BR. The evolution and maintenance of virulence in microparasites. Emerg Infect Dis. 1996;2(2):93-102. doi: 10.3201/eid0202.960203. PubMed PMID: 8903208; PubMed Central PMCID: PMCPMC2639826.

21. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, et al. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group Escherichia coli strains. Mol Biol Evol. 2007;24(11):2373-84. doi: 10.1093/molbev/msm172. PubMed PMID: 17709333.

22. Wold AE, Caugant DA, Lidin-Janson G, de Man P, Svanborg C. Resident colonic Escherichia coli strains frequently display uropathogenic characteristics. J Infect Dis. 1992;165(1):46-52. PubMed PMID: 1727897.

23. Norton JP, Mulvey MA. Toxin-antitoxin systems are important for niche-specific colonization and stress resistance of uropathogenic Escherichia coli. PLoS Pathog. 2012;8(10):e1002954. doi: 10.1371/journal.ppat.1002954. PubMed PMID: 23055930; PubMed Central PMCID: PMCPMC3464220.

24. Russell CW, Mulvey MA. The Extraintestinal Pathogenic Escherichia coli Factor RqlI Constrains the Genotoxic Effects of the RecQ-Like Helicase RqlH. PLoS Pathog. 2015;11(12):e1005317. doi: 10.1371/journal.ppat.1005317. PubMed PMID: 26636713; PubMed Central PMCID: PMCPMC4670107.

25. Myhal ML, Laux DC, Cohen PS. Relative colonizing abilities of human fecal and K 12 strains of Escherichia coli in the large intestines of streptomycin-treated mice. Eur J Clin Microbiol. 1982;1(3):186-92. PubMed PMID: 6756909.

26. Rosen MJ, Dhawan A, Saeed SA. Inflammatory Bowel Disease in Children and Adolescents. JAMA Pediatr. 2015;169(11):1053-60. doi: 10.1001/jamapediatrics.2015.1982. PubMed PMID: 26414706; PubMed Central PMCID: PMCPMC4702263.

27. Troeger H, Richter JF, Beutin L, Gunzel D, Dobrindt U, Epple HJ, et al. Escherichia coli alpha-haemolysin induces focal leaks in colonic epithelium: a novel mechanism of bacterial translocation. Cell Microbiol. 2007;9(10):2530-40. doi: 10.1111/j.1462-5822.2007.00978.x. PubMed PMID: 17587334.

28. Bucker R, Schulz E, Gunzel D, Bojarski C, Lee IF, John LJ, et al. alpha-Haemolysin of Escherichia coli in IBD: a potentiator of inflammatory activity in the colon. Gut. 2014;63(12):1893-901. doi: 10.1136/gutjnl-2013-306099. PubMed PMID: 24534723.

29. Welch RA. Uropathogenic Escherichia coli-Associated Exotoxins. Microbiol Spectr. 2016;4(3). doi: 10.1128/microbiolspec.UTI-0011-2012. PubMed PMID: 27337488; PubMed Central PMCID: PMCPMC4920080.

30. Fiorentini C, Arancia G, Caprioli A, Falbo V, Ruggeri FM, Donelli G. Cytoskeletal changes induced in HEp-2 cells by the cytotoxic necrotizing factor of Escherichia coli. Toxicon. 1988;26(11):1047-56. PubMed PMID: 3072687.

31. Caprioli A, Falbo V, Roda LG, Ruggeri FM, Zona C. Partial purification and characterization of an escherichia coli toxic factor that induces morphological cell alterations. Infect Immun. 1983;39(3):1300-6. PubMed PMID: 6341235; PubMed

Central PMCID: PMCPMC348099.

32. Smith YC, Rasmussen SB, Grande KK, Conran RM, O'Brien AD. Hemolysin of uropathogenic Escherichia coli evokes extensive shedding of the uroepithelium and hemorrhage in bladder tissue within the first 24 hours after intraurethral inoculation of mice. Infect Immun. 2008;76(7):2978-90. doi: 10.1128/IAI.00075-08. PubMed PMID: 18443089; PubMed Central PMCID: PMCPMC2446707.

33. Rippere-Lampe KE, Lang M, Ceri H, Olson M, Lockman HA, O'Brien AD. Cytotoxic necrotizing factor type 1-positive Escherichia coli causes increased inflammation and tissue damage to the prostate in a rat prostatitis model. Infect Immun. 2001;69(10):6515-9. doi: 10.1128/IAI.69.10.6515-6519.2001. PubMed PMID: 11553597; PubMed Central PMCID: PMCPMC98788.

34. Rippere-Lampe KE, O'Brien AD, Conran R, Lockman HA. Mutation of the gene encoding cytotoxic necrotizing factor type 1 (cnf(1)) attenuates the virulence of uropathogenic Escherichia coli. Infect Immun. 2001;69(6):3954-64. doi: 10.1128/IAI.69.6.3954-3964.2001. PubMed PMID: 11349064; PubMed Central PMCID: PMCPMC98434.

35. Johnson DE, Drachenberg C, Lockatell CV, Island MD, Warren JW, Donnenberg MS. The role of cytotoxic necrotizing factor-1 in colonization and tissue injury in a murine model of urinary tract infection. FEMS Immunol Med Microbiol. 2000;28(1):37-41. PubMed PMID: 10767605.

36. Kanamaru S, Kurazono H, Ishitoya S, Terai A, Habuchi T, Nakano M, et al. Distribution and genetic association of putative uropathogenic virulence factors iroN, iha, kpsMT, ompT and usp in Escherichia coli isolated from urinary tract infections in Japan. J Urol. 2003;170(6 Pt 1):2490-3. doi: 10.1097/01.ju.0000094185.48467.dc. PubMed PMID: 14634457.

37. Bauer RJ, Zhang L, Foxman B, Siitonen A, Jantunen ME, Saxen H, et al. Molecular epidemiology of 3 putative virulence genes for Escherichia coli urinary tract infection-usp, iha, and iroN(E. coli). J Infect Dis. 2002;185(10):1521-4. doi: 10.1086/340206. PubMed PMID: 11992291.

38. Ostblom A, Adlerberth I, Wold AE, Nowrouzian FL. Pathogenicity island markers, virulence determinants malX and usp, and the capacity of Escherichia coli to persist in infants' commensal microbiotas. Appl Environ Microbiol. 2011;77(7):2303-8. doi: 10.1128/AEM.02405-10. PubMed PMID: 21317254; PubMed Central PMCID: PMCPMC3067437.

39. Nipic D, Podlesek Z, Budic M, Crnigoj M, Zgur-Bertok D. Escherichia coli uropathogenic-specific protein, Usp, is a bacteriocin-like genotoxin. J Infect Dis. 2013;208(10):1545-52. doi: 10.1093/infdis/jit480. PubMed PMID: 23997234.

40. Zaw MT, Yamasaki E, Yamamoto S, Nair GB, Kawamoto K, Kurazono H. Uropathogenic specific protein gene, highly distributed in extraintestinal uropathogenic Escherichia coli, encodes a new member of H-N-H nuclease superfamily. Gut Pathog. 2013;5(1):13. doi: 10.1186/1757-4749-5-13. PubMed PMID: 23759109; PubMed Central PMCID: PMCPMC3685522.

41. Kurazono H, Yamamoto S, Nakano M, Nair GB, Terai A, Chaicumpa W, et al. Characterization of a putative virulence island in the chromosome of uropathogenic Escherichia coli possessing a gene encoding a uropathogenic-specific protein. Microb Pathog. 2000;28(3):183-9. doi: 10.1006/mpat.1999.0331. PubMed PMID: 10702359.

42. Nougayrede JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science. 2006;313(5788):848-51. doi: 10.1126/science.1127059. PubMed PMID: 16902142.

43. Cuevas-Ramos G, Petit CR, Marcq I, Boury M, Oswald E, Nougayrede JP. Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. Proc Natl Acad Sci U S A. 2010;107(25):11537-42. doi: 10.1073/pnas.1001261107. PubMed PMID: 20534522; PubMed Central PMCID: PMCPMC2895108.

44. Taieb F, Petit C, Nougayrede JP, Oswald E. The Enterobacterial Genotoxins: Cytolethal Distending Toxin and Colibactin. EcoSal Plus. 2016;7(1). doi: 10.1128/ecosalplus.ESP-0008-2016. PubMed PMID: 27419387.

45. McCarthy AJ, Martin P, Cloup E, Stabler RA, Oswald E, Taylor PW. The Genotoxin Colibactin Is a Determinant of Virulence in Escherichia coli K1 Experimental Neonatal Systemic Infection. Infect Immun. 2015;83(9):3704-11. doi: 10.1128/IAI.00716-15. PubMed PMID: 26150540; PubMed Central PMCID: PMCPMC4534652.

46. Marcq I, Martin P, Payros D, Cuevas-Ramos G, Boury M, Watrin C, et al. The genotoxin colibactin exacerbates lymphopenia and decreases survival rate in mice infected with septicemic Escherichia coli. J Infect Dis. 2014;210(2):285-94. doi: 10.1093/infdis/jiu071. PubMed PMID: 24489107.

47. Payros D, Secher T, Boury M, Brehin C, Menard S, Salvador-Cartier C, et al. Maternally acquired genotoxic Escherichia coli alters offspring's intestinal homeostasis. Gut Microbes. 2014;5(3):313-25. doi: 10.4161/gmic.28932. PubMed PMID: 24971581; PubMed Central PMCID: PMCPMC4153768.

48. Homburg S, Oswald E, Hacker J, Dobrindt U. Expression analysis of the colibactin gene cluster coding for a novel polyketide in Escherichia coli. FEMS Microbiol Lett. 2007;275(2):255-62. doi: 10.1111/j.1574-6968.2007.00889.x. PubMed PMID: 17714479.

49. Morimoto YV, Minamino T. Structure and function of the bi-directional bacterial flagellar motor. Biomolecules. 2014;4(1):217-34. doi: 10.3390/biom4010217. PubMed PMID: 24970213; PubMed Central PMCID: PMCPMC4030992.

50. Haugen BJ, Pellett S, Redford P, Hamilton HL, Roesch PL, Welch RA. In vivo gene expression analysis identifies genes required for enhanced colonization of the mouse urinary tract by uropathogenic Escherichia coli strain CFT073 dsdA. Infect Immun. 2007;75(1):278-89. doi: 10.1128/IAI.01319-06. PubMed PMID: 17074858; PubMed Central PMCID: PMCPMC1828413.

51. Wright KJ, Seed PC, Hultgren SJ. Uropathogenic Escherichia coli flagella aid in efficient urinary tract colonization. Infect Immun. 2005;73(11):7657-68. doi:

10.1128/IAI.73.11.7657-7668.2005. PubMed PMID: 16239570; PubMed Central PMCID: PMCPMC1273872.

52. Lane MC, Lockatell V, Monterosso G, Lamphier D, Weinert J, Hebel JR, et al. Role of motility in the colonization of uropathogenic Escherichia coli in the urinary tract. Infect Immun. 2005;73(11):7644-56. doi: 10.1128/IAI.73.11.7644-7656.2005. PubMed PMID: 16239569; PubMed Central PMCID: PMCPMC1273871.

53. Lane MC, Alteri CJ, Smith SN, Mobley HL. Expression of flagella is coincident with uropathogenic Escherichia coli ascension to the upper urinary tract. Proc Natl Acad Sci U S A. 2007;104(42):16669-74. doi: 10.1073/pnas.0607898104. PubMed PMID: 17925449; PubMed Central PMCID: PMCPMC2034267.

54. McCormick BA, Laux DC, Cohen PS. Neither motility nor chemotaxis plays a role in the ability of Escherichia coli F-18 to colonize the streptomycin-treated mouse large intestine. Infect Immun. 1990;58(9):2957-61. PubMed PMID: 2201640; PubMed Central PMCID: PMCPMC313594.

55. Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, Sabo A, et al. Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: a comparative genomics approach. Proc Natl Acad Sci U S A. 2006;103(15):5977-82. doi: 10.1073/pnas.0600938103. PubMed PMID: 16585510; PubMed Central PMCID: PMCPMC1424661.

56. Cusumano CK, Hung CS, Chen SL, Hultgren SJ. Virulence plasmid harbored by uropathogenic Escherichia coli functions in acute stages of pathogenesis. Infect Immun. 2010;78(4):1457-67. doi: 10.1128/IAI.01260-09. PubMed PMID: 20123719; PubMed Central PMCID: PMCPMC2849428.

57. Wijetunge DS, Karunathilake KH, Chaudhari A, Katani R, Dudley EG, Kapur V, et al. Complete nucleotide sequence of pRS218, a large virulence plasmid, that augments pathogenic potential of meningitis-associated Escherichia coli strain RS218. BMC Microbiol. 2014;14:203. doi: 10.1186/s12866-014-0203-9. PubMed PMID: 25164788; PubMed Central PMCID: PMCPMC4155114.

58. Russell CW, Mulvey MA. Type 1 and P pili of uropathogenic Escherichia coli. In: Barocchi MA, Telford JL, editors. Bacterial pili: structure, synthesis and role in disease. 1 ed. Wallingford, UK: CAB International; 2014. p. 49-70.

59. Lane MC, Simms AN, Mobley HL. complex interplay between type 1 fimbrial expression and flagellum-mediated motility of uropathogenic Escherichia coli. J Bacteriol. 2007;189(15):5523-33. doi: 10.1128/JB.00434-07. PubMed PMID: 17513470; PubMed Central PMCID: PMC1951814.

60. Simms AN, Mobley HL. Multiple genes repress motility in uropathogenic Escherichia coli constitutively expressing type 1 fimbriae. J Bacteriol. 2008;190(10):3747-56. doi: 10.1128/JB.01870-07. PubMed PMID: 18359812; PubMed Central PMCID: PMC2395009.

61. Cooper LA, Simmons LA, Mobley HL. Involvement of mismatch repair in the reciprocal control of motility and adherence of uropathogenic Escherichia coli. Infect Immun. 2012;80(6):1969-79. doi: 10.1128/IAI.00043-12. PubMed PMID: 22473602;

PubMed Central PMCID: PMC3370570.

62. Leatham MP, Stevenson SJ, Gauger EJ, Krogfelt KA, Lins JJ, Haddock TL, et al. Mouse intestine selects nonmotile flhDC mutants of Escherichia coli MG1655 with increased colonizing ability and better utilization of carbon sources. Infect Immun. 2005;73(12):8039-49. doi: 10.1128/IAI.73.12.8039-8049.2005. PubMed PMID: 16299298; PubMed Central PMCID: PMCPMC1307065.

63. Giraud A, Arous S, De Paepe M, Gaboriau-Routhiau V, Bambou JC, Rakotobe S, et al. Dissecting the genetic components of adaptation of Escherichia coli to the mouse gut. PLoS Genet. 2008;4(1):e2. doi: 10.1371/journal.pgen.0040002. PubMed PMID: 18193944; PubMed Central PMCID: PMCPMC2174974.

64. Vosti KL. Infections of the urinary tract in women: a prospective, longitudinal study of 235 women observed for 1-19 years. Medicine (Baltimore). 2002;81(5):369-87. PubMed PMID: 12352632.

65. Mulvey MA, Lopez-Boado YS, Wilson CL, Roth R, Parks WC, Heuser J, et al. Induction and Evasion of Host Defenses by Type 1-Piliated Uropathogenic Escherichia coli. Science. 1998;282(5393):1494-7. doi: 10.1126/science.282.5393.1494.

66. Adiba S, Nizak C, van Baalen M, Denamur E, Depaulis F. From grazing resistance to pathogenesis: the coincidental evolution of virulence factors. PLoS One. 2010;5(8):e11882. doi: 10.1371/journal.pone.0011882. PubMed PMID: 20711443; PubMed Central PMCID: PMC2920306.

67. Nickel JC. Practical management of recurrent urinary tract infections in premenopausal women. Rev Urol. 2005;7(1):11-7. PubMed PMID: 16985802; PubMed Central PMCID: PMCPMC1477561.

68. Wold AE, Mestecky J, Tomana M, Kobata A, Ohbayashi H, Endo T, et al. Secretory immunoglobulin A carries oligosaccharide receptors for Escherichia coli type 1 fimbrial lectin. Infection and immunity. 1990;58(9):3073-7. Epub 1990/09/01. PubMed PMID: 2201644; PubMed Central PMCID: PMC313613.

69. Bollinger RR, Everett ML, Wahl SD, Lee YH, Orndorff PE, Parker W. Secretory IgA and mucin-mediated biofilm formation by environmental strains of Escherichia coli: role of type 1 pili. Mol Immunol. 2006;43(4):378-87. doi: 10.1016/j.molimm.2005.02.013. PubMed PMID: 16310051.

70. Friman V, Adlerberth I, Connell H, Svanborg C, Hanson LA, Wold AE. Decreased expression of mannose-specific adhesins by Escherichia coli in the colonic microflora of immunoglobulin A-deficient individuals. Infect Immun. 1996;64(7):2794-8. PubMed PMID: 8698510; PubMed Central PMCID: PMCPMC174141.

71. Green SI, Ajami NJ, Ma L, Poole NM, Price RE, Petrosino JF, et al. Murine model of chemotherapy-induced extraintestinal pathogenic Escherichia coli translocation. Infect Immun. 2015;83(8):3243-56. doi: 10.1128/IAI.00684-15. PubMed PMID: 26034214; PubMed Central PMCID: PMCPMC4496622.

72. Wadolkowski EA, Laux DC, Cohen PS. Colonization of the streptomycin-treated

mouse large intestine by a human fecal Escherichia coli strain: role of growth in mucus. Infect Immun. 1988;56(5):1030-5. PubMed PMID: 3281898; PubMed Central PMCID: PMCPMC259757.

73. Anderson JD, Gillespie WA, Richmond MH. Chemotherapy and antibiotic-resistance transfer between Enterobacteria in the human gastro-intestinal tract. J Med Microbiol. 1973;6(4):461-73. doi: 10.1099/00222615-6-4-461. PubMed PMID: 4587394.

74. Lawley TD, Walker AW. Intestinal colonization resistance. Immunology. 2013;138(1):1-11. doi: 10.1111/j.1365-2567.2012.03616.x. PubMed PMID: 23240815; PubMed Central PMCID: PMCPMC3533696.

75. Rausch P, Basic M, Batra A, Bischoff SC, Blaut M, Clavel T, et al. Analysis of factors contributing to variation in the C57BL/6J fecal microbiota across German animal facilities. Int J Med Microbiol. 2016. doi: 10.1016/j.ijmm.2016.03.004. PubMed PMID: 27053239.

76. Murphy KC, Campellone KG. Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic E. coli. BMC Mol Biol. 2003;4:11. doi: 10.1186/1471-2199-4-11. PubMed PMID: 14672541; PubMed Central PMCID: PMCPMC317293.

77. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97(12):6640-5. doi: 10.1073/pnas.120163297. PubMed PMID: 10829079; PubMed Central PMCID: PMCPMC18686.

78. Li XT, Thomason LC, Sawitzke JA, Costantino N, Court DL. Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. Nucleic Acids Res. 2013;41(22):e204. doi: 10.1093/nar/gkt1075. PubMed PMID: 24203710; PubMed Central PMCID: PMCPMC3905872.

CHAPTER 6

DISCUSSION

The Beginning of the End

This dissertation has focused on the various factors that ExPEC requires to colonize the host, including adhesins, toxins, plasmids, metabolism, and protection of bacterial DNA, with a special emphasis on how ExPEC colonizes the gut. In addition to being a simple hunt for bacterial fitness factors, this work touches on broader matters, including how the gut can act as a reservoir for recurring infections, the difficulty of selecting an appropriate mouse model, and the evolution of pathogenesis, which are discussed below.

ExPEC Reservoirs for Recurrent UTIs

Although antibiotic treatment is effective for resolving a urinary tract infection in most patients, a major challenge for some patients is the issue of recurrence. One study that followed over 1,000 women for a decade found a wide range of UTI frequency, with 10.6% of patients experiencing no infection, to ~52% of patients suffering from 2-6 infections, and one patient with 42 UTI episodes [1]. A critical question surrounding recurrent UTIs (rUTIs) has been whether the ExPEC strain that causes the initial infection is the same strain responsible for subsequent infections. Unfortunately, the answer to that question appears to depend on the study [2]. However, it is clear that many patients are infected by a new strain at each episode, indicating that host susceptibility may be an important factor that determines rUTI frequency. In many other patients, the same strain can be isolated during multiple UTI episodes [2, 3], suggesting that in these cases, a reservoir of ExPEC exists either in the patient or within their environment that continues to supply bacteria even after the initial UTI episode is resolved.

There could be many sources of ExPEC in the case of patients that are infected with the same strain multiple times. A likely reservoir is the gut as it is home to relatively

high levels of ExPEC bacteria. This possibility is strengthened by an analysis of 6 patients that collectively experienced 13 rUTI episodes [3]. It was observed that the infecting strain could be detected in the feces between episodes in 12 (92%) of the cases. In 3 of these patients, antibiotic treatment only temporarily made the strain undetectable in the feces, whereas in the other 3 patients, the antibiotic treatment did not alter fecal levels of the pathogen at all. Therefore, it is possible that during antibiotic treatment for an ongoing UTI, the gut-localized population of the infecting strain is not successfully cleared, and that it is capable of reseeding the urinary tract for a subsequent infection.

In addition to the gut, there are other possible ExPEC reservoirs that could be driving rUTIs. For example, during bladder infection, ExPEC bacteria are capable of invading into the bladder epithelial cells and living intracellularly, sheltered from antibiotic treatment [4]. Within the superficial host cells, the bacteria quickly replicate and form intracellular bacterial communities, which can exit the host cell into the bladder lumen. Within deeper layers of the epithelium, the bacteria can remain quiescent for longer periods of time, but later resurge and potentially reinitiate symptoms. Other reservoirs are also possible, including within pets, a sexual partner, or the environment [5, 6].

The treatment of rUTIs is challenging, as a delicate balance must be maintained between helping the patient and avoiding selection for antibiotic resistance or disruption of the gut microbiota. For women with 2-4 UTIs per year, an economical and practical strategy is patient-initiated antimicrobial treatment [7]. This strategy allows the patient to keep antibiotics on hand and begin treatment immediately upon the appearance of symptoms without having to wait for an appointment and a prescription. For women with more frequent UTIs, a prophylactic approach can be taken through either continuous administration of low-dose antibiotics, or a single dose taken immediately after intercourse [7].

An attractive future treatment option for rUTI would be to remove the ExPEC reservoir in addition to treating the UTI. For example, combining antibiotic treatment with administration of chitosan—a chemical that facilitates exfoliation of the bladder epithelium—can reduce overall bacterial titers in infected mice, presumably by removing intracellular reservoirs [8]. A more difficult task would be to remove ExPEC from the gut reservoir without pathologic disruption of the microbiota. However, such a strategy could possibly reduce rUTIs in patients that are in desperate need of an effective treatment. The work presented in this dissertation could perhaps be used in the future to design such a treatment, as many possible therapeutic targets have been identified, including RqlI, GlpR, and type 1 pili. Unfortunately, like most antimicrobial treatments, it is fairly easy to imagine resistance evolving. For example, a simple loss-of-function mutation in *rqlH* would render an RqlI-targeted therapy useless. Due to the interconnectedness of metabolism, it is likely that there are mutations that would also nullify targeting the glycerol degradation pathway. In contrast, attacking type 1 pili with mannosides has already been shown to be effective at clearing ExPEC from the gut (Spaulding, et al., personal communication).

Limitations of the Mouse Model of Gut Colonization

One major factor that influences the outcome and interpretation of gut colonization experiments is the choice of mouse model. The most common protocol for *E. coli* gut colonization has been the streptomycin-treated mouse model, in which the antibiotic streptomycin is added to the water prior to inoculation, and is continuously administered throughout the experiment [9]. This model is convenient due to the ease of antibiotic administration and the consistency with which the mice are colonized. The major drawback for this method is the necessary disruption of the microbiota. Another system utilizes germ-free mice, which allows for easy colonization by *E. coli*, and which

has the advantage of simplifying interpretation of the results as the complexity of the microbiota is sidestepped. However, germ-free mice exhibit many phenotypes that differ from conventional mice [10].

In this work, we chose to colonize conventional mice without any antibiotic treatment, allowing us to observe gut colonization by ExPEC in the presence of a complete microbiota and a normal host environment. The mice were colonized at a high frequency, and the bacteria persisted for long periods of time. This is contrary to expectations put forth in studies utilizing the streptomycin-treated mouse model, a model that is based on the observation that antibiotic treatment is generally required to provide a niche for *E. coli* to inhabit. Furthermore, through personal communications, we have learned that other ExPEC researchers have failed to colonize conventional mice with ExPEC in the absence of streptomycin. There appears to be something unique about our mouse colony that allows the mice to be readily colonized by ExPEC. We hypothesize that the susceptibility of our mice to ExPEC colonization is likely due to differences in microbiota composition, as the microbiota is known to differ among animal facilities [11]. For example, we have never detected any *E. coli* in our mice prior to colonization by ExPEC, neither by plating feces onto MacConkey agar nor by 16S sequencing, whereas others have isolated endogenous ExPEC strains from their colonies [12]. The lack of endogenous *E. coli* may facilitate ExPEC gut colonization in our experiments, but there are still *E. coli* strains that do not colonize well in our model, such as MG1655, a non-pathogenic K12 derivative that is frequently used in gut colonization experiments. Therefore, this model of colonization allows us to test aspects of ExPEC gut colonization in the face of an undisturbed microbiota, with the drawback of not being easily repeated at other facilities. However, this limitation itself is interesting. Future work to define the differences in ExPEC gut persistence between facilities could be invaluable to understanding the process of colonization, and how this process is altered with different

microbiota compositions.

Another limitation of our *in vivo* model is the relatively high amount of deviation in the data, especially when two bacterial strains are competed. The "spread" in data points tends to increase over the course of the experiment and complicate interpretation. For example, when the Δ*rqlHI* strain was competed against the WT strain in the mouse gut, the standard deviation rose from 0.25 at day 5 to 1.49 at day 50 (Fig 3.8C). At day 50, the competitive index ranged from -3.34 (2,200-fold more WT than mutant) to 2.24 (170-fold more mutant than WT). However, the median competitive index steadily held close to 0, and most of the indices were smaller values between 1 and -1, making it easier to conclude that there was no difference between the WT and mutant strain. Another group made similar observations when competing two *E. coli* strains with equal fitness [13]. Two derivatives of MG1655 were created in which either YFP or CFP were expressed, but which were otherwise identical. When these two strains were competed in streptomycin-treated mice, the expectation that the ratio between the two strains would remain at 1:1 throughout the experiment did not hold true in many of the mice. In several mice, the YFP strain became a minor part of the total MG1655 population, while in other mice, it dominated by the end of the experiment. As several other studies have reported that *E. coli* can acquire consequential mutations during gut colonization experiments [14-18], the authors tested whether evolution was occurring in their study. Indeed, over the course of the experiment, the strains developed increased gut fitness, and were able to outcompete the ancestral strain in subsequent assays. By sequencing these derivatives, many loss-of-function mutations in the *gat* operon were observed, which would eliminate the ability of the strain to utilize galactitol, but are hypothesized to improve its ability to utilize other sugars. The alteration of metabolism is a common result of the short-term evolution occurring in these studies [15-17]. We believe that such evolution is occurring in our experiments, and is leading to increased noise in the data.

Unfortunately, various attempts to reduce the deviation in the data failed. Future work to define the beneficial mutations could allow us to mitigate the effects of this phenomenon and make the data easier to interpret.

The Evolution of Extraintestinal Pathogenesis

A major question that remains in ExPEC research is how extraintestinal pathogenesis came to be, and why it might be beneficial to the bacteria. To answer these questions, the selective pressures that ExPEC has experienced need to be defined. It is often assumed that ExPEC evolution has been mostly driven by the gut environment [19]. This assumption is derived from the argument that the gut is the major ExPEC niche, and that extraintestinal infections tend to be shorter in duration in comparison to the time that ExPEC spends within the gut. The shorter duration limits the transmission of selected clones to other hosts. At its extreme, infection of an extraintestinal niche can be a dead-end, as in the case of lethal bacteremia or meningitis where transmission from one host to another is unlikely. However, whether infection of the urinary tract is also a dead-end is debatable, especially when the pre-antibiotic era is considered. One study tracked nonpregnant women suffering from cystitis who received a placebo pill instead of an antibiotic [20]. A little over 70% of the women had cleared the infection spontaneously within 4 weeks. Although in most cases the infection was resolved relatively quickly, a substantial proportion (~30%) continued to exhibit symptoms and shed bacteria in their urine 1 month after infection. This suggests that prior to antibiotics (or even today in situations where antibiotics are not readily available), there existed a larger window of opportunity for ExPEC transmission from the urine of one individual to another individual or the environment. Although transmission via this route is substantially more modest than through the fecal route, and probably even less frequent now than in the past, it may not strictly be a dead-end. To my knowledge, a comparison

of the rate at which ExPEC is transmitted from one individual to another, or to the environment, via feces and urine has not been determined.

If one operates under the assumption that the gut is the major host-associated evolutionary arena for ExPEC—and that other niches, such as the urinary tract, play a minor role—then the mechanism driving evolution of extraintestinal virulence factors is a puzzle. How could a gene evolve to be used specifically in the blood, for example, when that niche is rarely encountered, and selected clones are passed on very infrequently? Three hypotheses explaining the evolution of virulence in general [21] can been considered in the context of ExPEC pathogenesis. First, the "conventional wisdom" hypothesis is that the coevolution of host and pathogen is always moving towards a commensal relationship. Therefore, a pathogenic interaction is a relatively young host-microbe association that will become less antagonistic over time. The drift from pathogen to commensal has been observed in various isolates such as Nissle 1917 and ABU 83972, which both exhibit degradation of pathogen-associated genes [22], but it is unclear whether this process is occurring for all ExPEC strains.

A second evolutionary idea is the trade-off hypothesis, which says that if the rate of transmission of a pathogen between individuals is higher than the rate of pathogen-mediated mortality, then virulence could be favored, especially if increased virulence translates to increased transmission [21]. One could imagine this balance coming into play during UTIs, in which the infection increases the frequency at which the individual urinates, which could possibly increase the transmission rate of the pathogen.

The third (and generally favored) evolutionary principle that is frequently applied to ExPEC is that of coincidental evolution. The idea is that extraintestinal virulence factors were actually shaped by selective pressures found in the gut, but that they happen to be useful in other niches, such as the urinary tract or blood [21]. Our data seem to indicate that this is not always true, as there are several canonical ExPEC virulence

factors that appear to play no role in bacterial fitness within the gut (see Chapter 5). Although it is difficult to extrapolate these findings to describe the entire evolutionary story of a pathogen, it is safe to say that evolutionary principles such as coincidental evolution may not apply to the entire genetic content of a bacterium. Rather, coincidental evolution could apply to some genes but not others. Furthermore, there are many other niches (e.g. the environment) and situations that need to be considered. For example, one study found a correlation between the ability of a given *E. coli* strain to resist being eating by amoeba and the extraintestinal virulence of that strain [23]. It may be that the ability of ExPEC to withstand attacks from amoeba both inside the gut and in the environment coincidentally equips it to be a pathogen in mammalian extraintestinal niches.

Concluding Remarks

The data presented in this dissertation begin to paint a picture of the required genes for ExPEC host colonization, especially within the gut. A lot of work is needed to better understand the colonization process, especially with an eye towards possible future therapies. The first major challenge in this endeavor is to simply identify genes important for gut colonization. The second—and possibly more difficult—task is to understand the functions of those genes. As a deeper knowledge of ExPEC host colonization is formed, therapeutic strategies to combat extraintestinal infections, rUTIs, and antibiotic resistance will improve.

References

1. Vosti KL. Infections of the urinary tract in women: a prospective, longitudinal study of 235 women observed for 1-19 years. Medicine (Baltimore). 2002;81(5):369-87. PubMed PMID: 12352632.

2. Ejrnaes K, Sandvang D, Lundgren B, Ferry S, Holm S, Monsen T, et al. Pulsed-field gel electrophoresis typing of Escherichia coli strains from samples collected before and after pivmecillinam or placebo treatment of uncomplicated community-acquired

urinary tract infection in women. J Clin Microbiol. 2006;44(5):1776-81. doi: 10.1128/JCM.44.5.1776-1781.2006. PubMed PMID: 16672406; PubMed Central PMCID: PMCPMC1479185.

3. Russo TA, Stapleton A, Wenderoth S, Hooton TM, Stamm WE. Chromosomal restriction fragment length polymorphism analysis of Escherichia coli strains causing recurrent urinary tract infections in young women. The Journal of infectious diseases. 1995;172(2):440-5. Epub 1995/08/01. PubMed PMID: 7622887.

4. Mulvey MA, Schilling JD, Hultgren SJ. Establishment of a persistent Escherichia coli reservoir during the acute phase of a bladder infection. Infection and immunity. 2001;69(7):4572-9. Epub 2001/06/13. doi: 10.1128/IAI.69.7.4572-4579.2001. PubMed PMID: 11402001; PubMed Central PMCID: PMC98534.

5. Brown PD, Foxman B. Pathogenesis of Urinary Tract Infection: the Role of Sexual Behavior and Sexual Transmission. Curr Infect Dis Rep. 2000;2(6):513-7. PubMed PMID: 11095901.

6. Johnson JR, Clabots C, Kuskowski MA. Multiple-host sharing, long-term persistence, and virulence of Escherichia coli clones from human and animal household members. J Clin Microbiol. 2008;46(12):4078-82. doi: 10.1128/JCM.00980-08. PubMed PMID: 18945846; PubMed Central PMCID: PMCPMC2593269.

7. Nickel JC. Practical management of recurrent urinary tract infections in premenopausal women. Rev Urol. 2005;7(1):11-7. PubMed PMID: 16985802; PubMed Central PMCID: PMCPMC1477561.

8. Blango MG, Ott EM, Erman A, Veranic P, Mulvey MA. Forced resurgence and targeting of intracellular uropathogenic Escherichia coli reservoirs. PloS one. 2014;9(3):e93327. doi: 10.1371/journal.pone.0093327. PubMed PMID: 24667805; PubMed Central PMCID: PMCPMC3965547.

9. Wadolkowski EA, Laux DC, Cohen PS. Colonization of the streptomycin-treated mouse large intestine by a human fecal Escherichia coli strain: role of growth in mucus. Infect Immun. 1988;56(5):1030-5. PubMed PMID: 3281898; PubMed Central PMCID: PMCPMC259757.

10. Umesaki Y. Use of gnotobiotic mice to identify and characterize key microbes responsible for the development of the intestinal immune system. Proceedings of the Japan Academy Series B, Physical and biological sciences. 2014;90(9):313-32. PubMed PMID: 25391317; PubMed Central PMCID: PMC4324924.

11. Rausch P, Basic M, Batra A, Bischoff SC, Blaut M, Clavel T, et al. Analysis of factors contributing to variation in the C57BL/6J fecal microbiota across German animal facilities. Int J Med Microbiol. 2016. doi: 10.1016/j.ijmm.2016.03.004. PubMed PMID: 27053239.

12. Lasaro M, Liu Z, Bishar R, Kelly K, Chattopadhyay S, Paul S, et al. Escherichia coli isolate for studying colonization of the mouse intestine and its application to two-component signaling knockouts. J Bacteriol. 2014;196(9):1723-32. doi: 10.1128/JB.01296-13. PubMed PMID: 24563035; PubMed Central PMCID: PMC3993324.

13. Barroso-Batista J, Sousa A, Lourenco M, Bergman ML, Sobral D, Demengeot J, et al. The first steps of adaptation of Escherichia coli to the gut are dominated by soft sweeps. PLoS Genet. 2014;10(3):e1004182. doi: 10.1371/journal.pgen.1004182. PubMed PMID: 24603313; PubMed Central PMCID: PMCPMC3945185.

14. Giraud A, Arous S, De Paepe M, Gaboriau-Routhiau V, Bambou JC, Rakotobe S, et al. Dissecting the genetic components of adaptation of Escherichia coli to the mouse gut. PLoS Genet. 2008;4(1):e2. doi: 10.1371/journal.pgen.0040002. PubMed PMID: 18193944; PubMed Central PMCID: PMC2174974.

15. Lescat M, Launay A, Ghalayini M, Magnan M, Glodt J, Pintard C, et al. Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an Escherichia coli natural isolate in the streptomycin treated mouse gut. Mol Ecol. 2016. doi: 10.1111/mec.13851. PubMed PMID: 27661780.

16. Leatham MP, Stevenson SJ, Gauger EJ, Krogfelt KA, Lins JJ, Haddock TL, et al. Mouse intestine selects nonmotile flhDC mutants of Escherichia coli MG1655 with increased colonizing ability and better utilization of carbon sources. Infect Immun. 2005;73(12):8039-49. doi: 10.1128/IAI.73.12.8039-8049.2005. PubMed PMID: 16299298; PubMed Central PMCID: PMCPMC1307065.

17. Leatham-Jensen MP, Frimodt-Moller J, Adediran J, Mokszycki ME, Banner ME, Caughron JE, et al. The streptomycin-treated mouse intestine selects Escherichia coli envZ missense mutants that interact with dense and diverse intestinal microbiota. Infect Immun. 2012;80(5):1716-27. doi: 10.1128/IAI.06193-11. PubMed PMID: 22392928; PubMed Central PMCID: PMCPMC3347456.

18. Lee SM, Wyse A, Lesher A, Everett ML, Lou L, Holzknecht ZE, et al. Adaptation in a mouse colony monoassociated with Escherichia coli K-12 for more than 1,000 days. Appl Environ Microbiol. 2010;76(14):4655-63. doi: 10.1128/AEM.00358-10. PubMed PMID: 20472724; PubMed Central PMCID: PMCPMC2901717.

19. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, et al. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group Escherichia coli strains. Mol Biol Evol. 2007;24(11):2373-84. doi: 10.1093/molbev/msm172. PubMed PMID: 17709333.

20. Mabeck CE. Treatment of uncomplicated urinary tract infection in non-pregnant women. Postgraduate medical journal. 1972;48(556):69-75. PubMed PMID: 4552445; PubMed Central PMCID: PMC2495172.

21. Levin BR. The evolution and maintenance of virulence in microparasites. Emerg Infect Dis. 1996;2(2):93-102. doi: 10.3201/eid0202.960203. PubMed PMID: 8903208; PubMed Central PMCID: PMCPMC2639826.

22. Vejborg RM, Friis C, Hancock V, Schembri MA, Klemm P. A virulent parent with probiotic progeny: comparative genomics of Escherichia coli strains CFT073, Nissle 1917 and ABU 83972. Molecular genetics and genomics : MGG. 2010;283(5):469-84. doi: 10.1007/s00438-010-0532-9. PubMed PMID: 20354866.

23. Adiba S, Nizak C, van Baalen M, Denamur E, Depaulis F. From grazing resistance to pathogenesis: the coincidental evolution of virulence factors. PloS one.

2010;5(8):e11882. doi: 10.1371/journal.pone.0011882. PubMed PMID: 20711443; PubMed Central PMCID: PMCPMC2920306.

APPENDIX

COMBINING QUANTITATIVE GENETIC FOOTPRINTING AND TRAIT ENRICHMENT ANALYSIS TO IDENTIFY FITNESS DETERMINANTS OF A BACTERIAL PATHOGEN

Reprinted with permission from Public Library of Science

Wiles TJ, Norton JP, Russell CW, Dalley BK, Fischer KF, Mulvey MA. (2013) Combining quantitative genetic footprinting and trait enrichment analysis to identify fitness determinants of a bacterial pathogen. PLoS Genet. 9(8):e1003716. doi: 10.1371/journal.pgen.1003716.

OPEN ACCESS Freely available online

PLOS | GENETICS

Combining Quantitative Genetic Footprinting and Trait Enrichment Analysis to Identify Fitness Determinants of a Bacterial Pathogen

Travis J. Wiles[1], J. Paul Norton[1], Colin W. Russell[1], Brian K. Dalley[2], Kael F. Fischer[1,3], Matthew A. Mulvey[1]*

1 Division of Microbiology and Immunology, Pathology Department, University of Utah School of Medicine, Salt Lake City, Utah, United States of America, **2** Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah, United States of America, **3** ARUP Laboratories, Salt Lake City, Utah, United States of America

Abstract

Strains of Extraintestinal Pathogenic *Escherichia coli* (ExPEC) exhibit an array of virulence strategies and are a major cause of urinary tract infections, sepsis and meningitis. Efforts to understand ExPEC pathogenesis are challenged by the high degree of genetic and phenotypic variation that exists among isolates. Determining which virulence traits are widespread and which are strain-specific will greatly benefit the design of more effective therapies. Towards this goal, we utilized a quantitative genetic footprinting technique known as transposon insertion sequencing (Tn-seq) in conjunction with comparative pathogenomics to functionally dissect the genetic repertoire of a reference ExPEC isolate. Using Tn-seq and high-throughput zebrafish infection models, we tracked changes in the abundance of ExPEC variants within saturated transposon mutant libraries following selection within distinct host niches. Nine hundred and seventy bacterial genes (18% of the genome) were found to promote pathogen fitness in either a niche-dependent or independent manner. To identify genes with the highest therapeutic and diagnostic potential, a novel Trait Enrichment Analysis (TEA) algorithm was developed to ascertain the phylogenetic distribution of candidate genes. TEA revealed that a significant portion of the 970 genes identified by Tn-seq have homologues more often contained within the genomes of ExPEC and other known pathogens, which, as suggested by the first axiom of molecular Koch's postulates, is considered to be a key feature of true virulence determinants. Three of these Tn-seq-derived pathogen-associated genes—a transcriptional repressor, a putative metalloendopeptidase toxin and a hypothetical DNA binding protein—were deleted and shown to independently affect ExPEC fitness in zebrafish and mouse models of infection. Together, the approaches and observations reported herein provide a resource for future pathogenomics-based research and highlight the diversity of factors required by a single ExPEC isolate to survive within varying host environments.

Citation: Wiles TJ, Norton JP, Russell CW, Dalley BK, Fischer KF, et al. (2013) Combining Quantitative Genetic Footprinting and Trait Enrichment Analysis to Identify Fitness Determinants of a Bacterial Pathogen. PLoS Genet 9(8): e1003716. doi:10.1371/journal.pgen.1003716

Editor: Danielle A. Garsin, The University of Texas Health Science Center at Houston, United States of America

Received December 13, 2012; **Accepted** June 26, 2013; **Published** August 22, 2013

Funding: This work was supported by grants AI095647, AI090369, and AI088086 from the National Institute of Allergy and Infectious Diseases. TJW and JPN were supported by NIH Genetics Training Grant T32-GM007464. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mulvey@path.utah.edu

Introduction

Within the *Escherichia coli* lineage there are several distinct virulent subgroups that are principally classified by an ability to cause a common set of diseases. One specific subgroup, Extraintestinal Pathogenic *E. coli* (ExPEC), is typically thought to be a benign inhabitant of the lower intestinal tract of warm-blooded vertebrates. However, outside this niche, ExPEC strains have the ability to persist in an array of secondary host-associated habitats where they can cause urinary tract infections (UTIs), meningitis, bacteremia and sepsis in both humans and domesticated animals [1,2,3,4]. The combined medical, agricultural and economic burden of ExPEC-related diseases is likely to increase as antibiotic resistance spreads [1,5,6]. Delineation of the genetic elements utilized by ExPEC to infect such a diverse spectrum of host niches and cause disease promises to advance our understanding of pathogen evolution and behavior while also highlighting more effective strategies to combat these pervasive opportunistic pathogens.

Despite colonizing similar ecological niches, ExPEC isolates can differ by 20–30% of their respective gene inventories and to date, a specific and ubiquitous molecular trait or function exclusive to this cohort has not been characterized [7,8]. Without knowledge of a unifying ExPEC-associated feature, the development of broad-spectrum therapeutic strategies is made exceptionally difficult [9]. Previous work indicates that individual ExPEC strains appear to have evolved distinct genetic repertoires that promote unique and at times subtle fitness advantages during colonization of specific host environments [10,11]. It is becoming clear that the working definition of ExPEC is more multigenic in nature. Further complicating the search for ExPEC-associated traits is the 30–40% of genes within their genomes that still require functional annotation [7]. This problem will likely persist as genome sequencing continues to outstrip rates of experimental characterization. Consequently, there is a pressing need to develop methods that mesh high-throughput genomics with context-based functional observations—a notion that is becoming increasingly appreciated [12,13]. To this end, we adapted a previously described

Author Summary

Antibiotic resistance is an increasingly serious problem, especially among pathogenic strains of *Escherichia coli* that cause urinary tract infections, sepsis and meningitis. It is important to obtain a more comprehensive genome-wide understanding of bacterial virulence because it has the potential to uncover novel and alternative therapeutic targets. Therefore, we probed the genome of a pathogenic *E. coli* isolate using transposon mutagenesis, deep sequencing and comparative pathogenomics in an effort to define its virulence gene repertoire. Using this multi-layered approach in combination with high-throughput zebrafish infection models, we identified hundreds of genes that affect pathogen fitness during localized and/or blood-borne infections. We also developed a bioinformatics-based method to systematically sift through our datasets for genes that are broadly conserved among an assortment of pathogenic species. Follow-up analysis of several pathogen-associated candidate genes using zebrafish and mouse infection models highlighted the capacity of our approach to identify novel fitness determinants. The results from this study are available via an interactive online data viewer (http://pathogenomics.path.utah.edu/F11_TnSeq/) so that investigators can more effectively search and utilize these findings.

transposon mutagenesis technique known as 'Tn-seq' to identify genes that promote pathogen fitness during colonization of a vertebrate host [14,15].

The use of transposons to conduct unbiased forward genetic screens in bacteria has provided many avenues for investigation over the last two decades [16,17,18]. However, these approaches are, for the most part, time and labor intensive and often lack reliable quantitative metrics and sampling depth to identify genes of interest for retrospective follow-up. On the other hand, transposon mutagenesis as a tool is quite amenable to innovation [19,20,21]. Tn-seq, also known as 'INSeq', utilizes deep sequencing to monitor the composition of insertion variants within bulk mutant pools [14,15]. This is accomplished through use of a modified *mariner* transposon that contains recognition sites within the two distal inverted repeats that are specific for the restriction enzyme MmeI. Cleavage of DNA by this enzyme occurs 18–20 base pairs from its recognition site, such that excision of the transposon by MmeI captures flanking genomic sequences. These sequences serve as tags that can be used to track the relative abundance of distinct insertion variants within heterogenetic mutant libraries. This technique ultimately allows the fitness of thousands of mutant variants to be quantified simultaneously following exposure to specific selective pressures. By comparing the composition and abundance of mutants before and after passage through selective conditions, such as within a host organism, a genome-wide map of loci that are important for bacterial fitness can be assembled.

The screening approach reported here involved injection of 48 h post-fertilization (hpf) zebrafish embryos with transposon-mutagenized pools of ExPEC derived from a single parent isolate. Embryonic zebrafish are particularly useful for this type of selection screen because their vertebrate physiology closely matches that of humans, and they have proven useful for the identification and characterization of virulence determinants that enable ExPEC to survive within host niches [10,11,22,23]. Of particular importance, zebrafish embryos possess many of the same innate defenses that mammalian hosts rely upon to resist ExPEC (e.g. phagocytes, complement and antimicrobial peptides) [10,24,25,26]. This attribute underscores the ability of the zebrafish host to mirror a set of overlapping selective pressures that ExPEC naturally encounter. Zebrafish embryos can be used to model both localized and systemic infections and the complete recovery and enumeration of bacteria from within the entire host is relatively straightforward [10,11]. Through three independent biological experiments employing Tn-seq we identified 970 genes (~18% of gene content) that promote the competitive fitness of ExPEC during either localized or systemic infections. A large number of these genes have unknown functions, while many others have reported roles in iron transport, bacterial secretion, two-component signaling and metabolism. To focus on genes that are likely important to ExPEC pathogenesis specifically, we devised a gene ontology-like method called 'TEA' (**T**rait **E**nrichment **A**nalysis), that enabled us to organize candidate genes according to their association with certain bacterial lineages or groups of bacteria that are phenotypically similar. The data presented in this report indicate that TEA, in combination with Tn-seq, provides an effective, streamlined approach for identifying biologically relevant fitness and virulence determinants that are employed by ExPEC within various host environments. Datasets generated in this study have been made freely available through a curated and searchable web-based data viewer (http://pathogenomics.path.utah.edu/F11_TnSeq/).

Results

Mutant library construction and screen design

We previously described the use of the embryonic zebrafish as a surrogate host to study genotype-phenotype relationships of a variety of human and non-human ExPEC isolates [10,11]. During this initial work we observed that the human cystitis (bladder infection) isolate F11 is particularly adept at growing within and eliciting death of zebrafish during either localized or systemic infections. Elucidation of the virulence gene repertoire (i.e. genetic determinants that promote pathogenic behaviors such as colonization or destruction of host tissues) of this ExPEC strain using Tn-seq provides a starting point for future comparative pathogenomics studies. Figure 1 presents an outline detailing the negative selection screen carried out in this study.

Transposon mutagenesis was accomplished by using the previously described pSAM vector and conjugation [15]. This plasmid, which contains a *mariner* transposon flanked by MmeI modified inverted repeats and the *himar*1C9 transposase, was retrofitted with an *E. coli* compatible antibiotic resistance cassette and promoter elements to generate pSAM-Ec (Figure S1). Multiple, independent mating events were performed to assemble three transposon mutant libraries to be used in three separate replicate screens (Figure 1A). It was determined that each pool contained a minimum of 50,000 distinct insertion variants by enumerating selectable, transposon positive colonies immediately following conjugation when mutant siblings are minimal. Successful transposition of single inserts into individual bacterial chromosomes was verified by detection of the transposon by Southern blot analysis (Figure S2). Because the cost of a pre-screen sequence analysis was prohibitive, we utilized a more economical means to confirm the diversity and saturation of mutant pools. The frequency of variants defective for lactose utilization, curli production and glycogen storage was determined using standard colorimetric plating assays (Figure S2), which demonstrated that our transposon mutant libraries were sufficiently saturated and ready for *in vivo* screening. Post-hoc sequencing later confirmed that our libraries contained 60,000 to 80,000 mutant variants each, corresponding to an insertion event approximately every 75 bp within the F11 chromosome (Table S1).

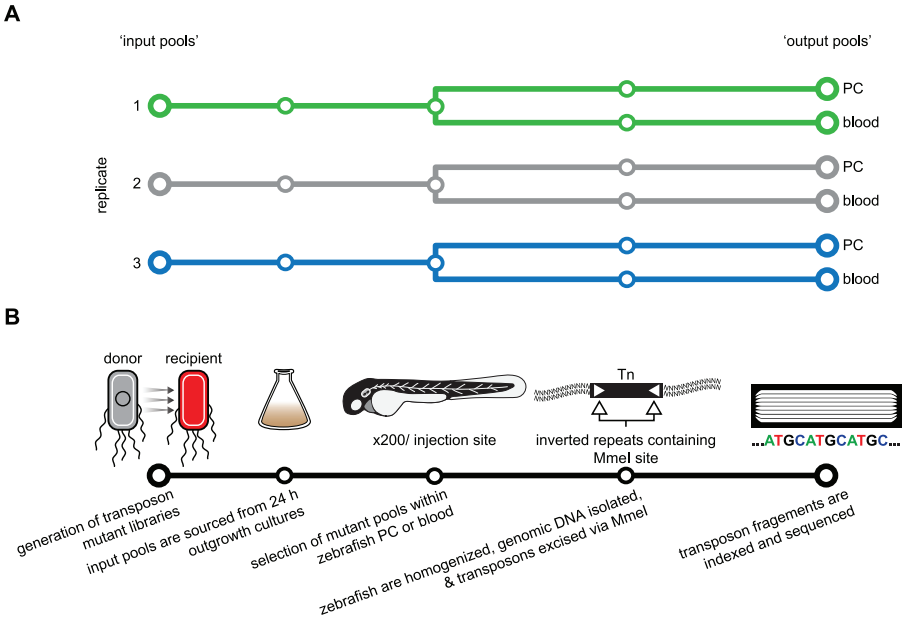


Figure 1. Schematic of the Tn-seq selection screen used to identify genes necessary for fitness in a zebrafish infection model. (A) Three selection screens were performed (green, grey and blue tracks) using three 'input' pools to generate a total of six 'output' pools. (B) For each replicate screen, an independent transposon mutant library was created. Expansion of mutant libraries from frozen stocks was done at 37°C for 24 h to produce input pools and starting inoculum. Approximately 200 zebrafish embryos were injected in the pericardial cavity (PC) or circulation valley (CV, blood) with approximately 3,000 mutant bacteria. After a selection period of 20 h, zebrafish were homogenized to facilitate recovery and extraction of total DNA. The restriction enzyme MmeI was used to excise transposons and flanking F11 genomic DNA for analysis by deep sequencing.
doi:10.1371/journal.pgen.1003716.g001

For each independent biological experiment, a single mutant library was cultivated from a frozen stock in M9 minimal media overnight at 37°C (Figure 1B). During this initial outgrowth period, variants with insertions that disabled genes critical for normal replication within broth culture were likely reduced in abundance or completely eliminated. Prior to injection of 48 hpf zebrafish embryos, a 1 ml aliquot of each overnight culture was pelleted, washed and suspended in phosphate buffered saline (PBS) at the desired density. A second aliquot was pelleted and stored as a reference 'input' population for later sequencing and comparative analyses with 'output' populations. A dose of approximately 3,000 colony forming units (CFU) from each mutant library was injected into either the pericardial cavity (PC) or into the blood via the circulation valley (CV) of zebrafish embryos (200 embryos per niche per experiment were used to ensure adequate sampling of mutant variants) (Figure S3). Injection of the PC provides a model of localized tissue infection, whereas injection into the CV results in rapid dissemination of ExPEC via the bloodstream, modeling bacteremia or sepsis-like infections [10,11]. We previously determined that a variety of ExPEC isolates have the capacity to grow unchecked within the PC, while only a subset are able to persist and multiply within the bloodstream. Presumably, the microenvironments encountered by ExPEC following injection into the zebrafish circulatory system are more challenging than the PC with respect to nutrient availability, host defenses or other as-yet-undefined factors [10,11]. F11 is able to survive and replicate within both the PC and blood. Once injected, selection proceeded for 18 to 20 h—a timeframe in which the mutant F11 population grows to between 10^4 and 10^6 CFU/embryo and elicits death in ~40–60% of the animals (Figure S4). Total genomic DNA of the surviving F11 variants was recovered from batch-homogenized fish (alive and dead) to produce output pools. In parallel, equal amounts of genomic material from input and corresponding output pools were digested using the MmeI restriction enzyme followed by enrichment of transposon-containing fragments with flanking genomic sequences via gel electrophoresis. The fragments were then excised and prepared for indexed sequencing on a single lane of an Illumina HiSeq 2000. General sequence-based features of each mutant pool, including number of detected mutant variants, sequencing depth, saturation and insertional bias, are summarized in Table S1.

Genetic footprinting by Tn-seq reveals distinguishable gene sets that have distinct functional compositions within the ExPEC genome

Gene insertion tolerance after passage in broth culture. Beyond targeting T+A dinucleotides, transposable

mariner elements have a relatively low insertional bias (Table S1 and Figure 2: F11 genome) [27]. Using sequence data from the three input samples exclusively, the number of insertion sequences mapped per 100 bp for all annotated regions was calculated to determine an insertion frequency (Table S2, Methods). As anticipated, we observed that transposon insertion rates within our mutant libraries mostly occurred at a steady rate independent of chromosomal region. An approach akin to traditional genetic footprinting was used to determine the specific insertional tolerance of each gene within the F11 genome [28,29]. Gene insertion rates that were more than 1 standard deviation below the genome-wide mean were considered hypo-tolerant (<2.1 inserts/100 bp, $n = 609$, Figure 2: hypo-tolerant), those within +/−1 standard deviation were tolerant (2.1 < inserts/100 bp <4.5, $n = 4,272$) and those that were more than 1 standard deviation above the genome-wide mean were hyper-tolerant (>4.5 inserts/100 bp, $n = 431$, Figure 2: hyper-tolerant). Genes assigned to the hypo-tolerant gene set presumably provide 'essential' or 'core' physiological functions. In contrast, genes within the hyper-tolerant gene set are likely not required during growth in broth culture and/or exhibit certain genetic features that support high rates of transposon insertion. Using the Kyoto Encyclopedia of Genes and Genomes (KEGG) functional hierarchies, the hypo-tolerant gene set, as expected, was found to be significantly enriched for informational functions related to transcription and translation and included various noncoding RNAs (ncRNA) and transfer RNAs (tRNA) (Figure 3A and B). The hyper-tolerant gene set contained a significantly elevated proportion of ribosomal RNAs, ncRNAs and genes annotated as 'hypothetical' (Figure 3B and F).

Genes required for *in vivo* fitness. To identify genes necessary for *in vivo* fitness, a modified genetic footprinting approach was used (Figure S5). For input-output pairs in each experimental replicate (i.e. insertion mutants present in cognate input and output pools), a normalized occurrence frequency reflecting their abundance was calculated. This approach relies on an ability to count the number of specific mutant variants in corresponding input and output pools. However, strong negative selection within the zebrafish host may have completely purged variants with disruptions in genes critical for fitness. Therefore, we established criteria to determine which mutant variants were likely lost due to selection and not by chance through experimental bottlenecks (Figure S5 and Methods). Missing variants that met these criteria were assigned a value equal to the limit of detection. Within each replicate experiment, we employed a Wilcoxon signed-rank test on the collection of inserts for each gene to determine if the distribution of insert occurrence frequencies between input and output samples was significantly different. This is a conservative approach that requires consistency between insertion variants across a given gene for it to be identified as a fitness determinant. Genes with 10 or more measured inserts, a Wilcoxon p value less than 0.05 and an average decrease in variant abundance of at least 2-fold were considered to have a significant influence on *in vivo* fitness (FDR <10%). Of the 970 genes identified across all replicates and niches, only 21 overlapped with the hypo-tolerant gene set, suggesting that the vast majority of genes identified are specifically required by F11 to survive within the zebrafish embryo (Figure 2: *in piscis*, http://pathogenomics.path.utah.edu/F11_TnSeq/). This '*in piscis*' gene set represents ~18% of the annotated regions within the F11 chromosome and is enriched for genes with functions that fall under the KEGG category 'environmental interactions' and the subcategories membrane and iron transport (Figure 3C). Additionally, metabolic functions were also highly represented, encompassing genes involved in glycan biosynthesis and the metabolism of amino acids and carbohydrates (Figure 3E). The proportion of candidate fitness genes that were identified within each of the three mutant libraries is visualized in Figure S6. Genes within the six different *in*

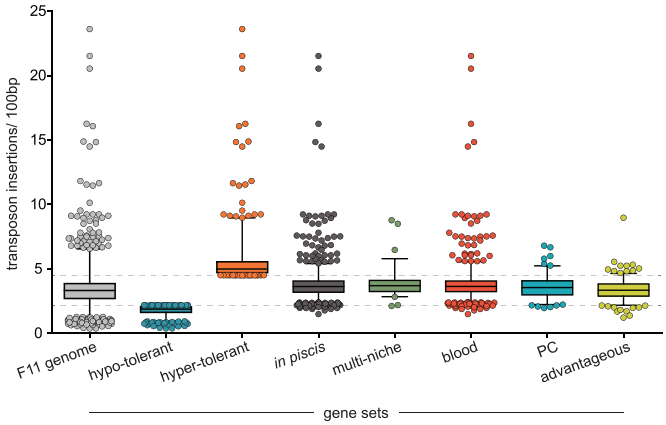


Figure 2. Transposon insertion rates for loci within different gene sets. From each input pool (1, 2 and 3) an average number of insertions detected per 100 bp was calculated for each annotated region within the F11 genome. Box plots indicate median and interquartile ranges with whiskers extending to the 1st and 99th percentiles. Dashed lines mark the upper and lower boundaries for hypo and hyper-tolerant gene sets, respectively. Gene set sizes: F11 genome - 5,312; hypo-tolerant - 609; hyper-tolerant - 431; *in piscis* - 970; multi-niche - 76; blood - 772; PC - 122; advantageous - 227. Note, the insertion rates reported here are expected to be overestimated—please refer to Methods section '*Candidate Identification*' for a complete description of how the number of transposon insertion events was determined.
doi:10.1371/journal.pgen.1003716.g002

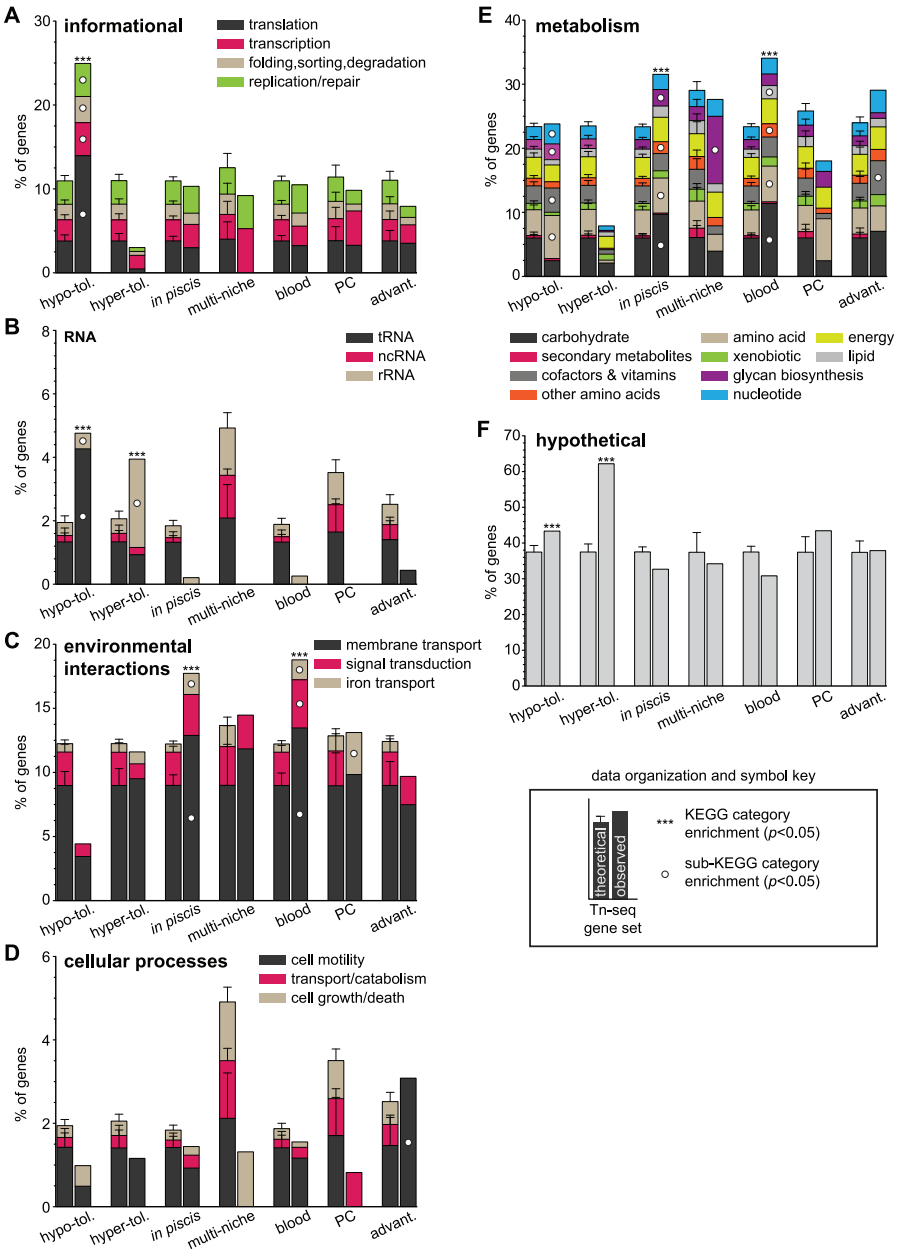

A
informational
translation
transcription
folding,sorting,degradation
replication/repair
% of genes
B
RNA
tRNA
ncRNA
rRNA
C
environmental interactions
membrane transport
signal transduction
iron transport
D
cellular processes
cell motility
transport/catabolism
cell growth/death
E
metabolism
carbohydrate
amino acid
energy
secondary metabolites
xenobiotic
lipid
cofactors & vitamins
glycan biosynthesis
other amino acids
nucleotide
F
hypothetical
hypo-tol.
hyper-tol.
in piscis
multi-niche
blood
PC
advant.
data organization and symbol key
theoretical
observed
Tn-seq gene set
KEGG category enrichment (p<0.05)
sub-KEGG category enrichment (p<0.05)

Figure 3. Functional KEGG categories enriched within gene sets. To orient the reader to graph organization and symbols, a key is provided in the lower right corner. The cumulative percent of genes (y-axis) within each Tn-seq-derived gene set (x-axis) belonging to the functional categories (A) informational, (B) RNA (C) environmental interactions, (D) cellular processes, (E) metabolism and (F) hypothetical is depicted. The proportion of genes contributed by specific sub-KEGG categories is represented by stacked elements within each column. Significant enrichments for KEGG and sub-KEGG categories are denoted with asterisks and white circles, respectively. Gene set abbreviations: hypo-tol. = hypo-tolerant; hyper-tol. = hyper-tolerant; advant. = advantageous. Theoretical KEGG category compositions for Tn-seq gene sets and *p* values were determined using 10,000 Monte Carlo simulations (Methods).
doi:10.1371/journal.pgen.1003716.g003

piscis gene sets showed only weak, though still in some cases significant, overlap (denoted in Figure S6), suggesting that the screening of multiple libraries was important for a more thorough enumeration of regions that influence *in vivo* fitness. Of note, we speculate that we did not fully reach a level of experimental saturation that matched the mutagenic saturation of our libraries. This observation may indicate a point of consideration for future Tn-seq-based screens using high-throughput surrogate hosts such as zebrafish where initial inoculum doses are relatively small and the stochastic nature of infection can influence the resolution of more subtle phenotypes. However, based on the findings presented in the following sections, we conclude that this perceived experimental noise did not compromise our ability to accurately identify and confirm genes of importance to ExPEC fitness and virulence.

Niche-dependent gene requirements. Seventy-six genes (1.4% of total gene content) were found to be necessary for fitness during both localized (PC) and systemic (blood) infections (http://pathogenomics.path.utah.edu/F11_TnSeq/). This 'multi-niche' gene set appears to collectively encode functions that directly influence gene expression and membrane architecture, both of which are known to be critical for pathogen fitness. Indeed, a number of the multi-niche genes have already been recognized for their role in bacterial stress resistance, regulation of virulence genes and pathogenesis (Table 1). Transposon inserts within several of the multi-niche genes resulted in some of the most dramatic declines in F11 fitness *in vivo* (Table 1). Notable candidates include *EcF11_3256/bipA*, *EcF11_1255/lrhA* and *EcF11_2271/rfaH* [30,31,32]. This gene set is also remarkably enriched for factors that function in glycan biosynthesis, including several genes involved in the production of capsular polysaccharides, which is a well-documented ExPEC-associated virulence trait (Figure 3E) [33,34].

As noted above, the zebrafish PC is more permissive to ExPEC growth than the bloodstream [10,11]. Corroborating this observation, we found 772 genes to be exclusively required for blood-borne fitness while only 122 genes were exclusive to the PC (http://pathogenomics.path.utah.edu/F11_TnSeq/). This blood-specific gene set was enriched for 7 different KEGG functional categories, whereas the PC-specific gene set was enriched for only 1 (Figure 3, $p<0.05$). Iron transport, the sole category enriched within the PC gene set, was also enriched within the blood set (Figure 3C). Additional categories that were enriched among blood-borne fitness genes covered signal transduction, membrane transport and the metabolism of amino acids, carbohydrates and lipids (Figure 3C and E). These observations indicate that the ability of ExPEC to survive and grow systemically within the zebrafish is, in part, dictated by a functionally diverse set of genes that facilitate the utilization of nutrients that are likely scavenged from the host.

Retrospective validation of a Tn-seq-derived fitness determinant. We chose the candidate gene *EcF11_3256/bipA*—which has not been previously reported to play a part in the virulence potential of ExPEC—to experimentally confirm the proficiency of our Tn-seq approach. An insertion map of this gene, depicting the reduction in fitness for each variant across all six pools, is presented in Figure 4A. The average changes in fitness for *bipA* and neighboring genes were also examined and indicated that the disruption of only *bipA* correlated with the observed fitness defect (Figure 4B). BipA is conserved across several bacterial phyla and has previously been shown to be a ribosome-associated GTPase that can coordinate the expression of virulence and stress response genes [30,35,36]. A targeted deletion of the *EcF11_3256/bipA* allele was generated and the resulting mutant derivative injected into the zebrafish PC and CV at a one-to-one ratio with wild type F11. F11Δ*3256* exhibited a significant reduction in fitness compared to the wild type strain in both environments (Figure 4C). Independent challenges within the PC and bloodstream revealed that F11Δ*3256* also exhibits significantly attenuated killing kinetics compared to wild type (Figure 4D). These results confirm that *EcF11_3256/bipA* is a critical component of F11's genetic repertoire and promotes both the fitness and virulence of this ExPEC strain within distinct host environments. Collectively, the identification of genes with known roles in bacterial pathogenesis (Table 1) and our observations with the Tn-seq-derived *in piscis* candidate gene *EcF11_3256/bipA* validate the Tn-seq protocol described here as a viable approach for detecting fitness and virulence determinants of ExPEC.

Disruptions that confer an *in vivo* fitness advantage. We identified 227 genes (~4% of gene content), that when disrupted conferred a fitness advantage to F11 within the zebrafish host (http://pathogenomics.path.utah.edu/F11_TnSeq/). Twenty of these gene disruptions were advantageous in both the PC and blood, 124 were advantageous only in the PC and 84 conveyed significant advantages in only the blood. KEGG functional categories enriched within the 'advantageous' gene set include cell motility, nucleotide metabolism and biosynthesis of vitamins and cofactors (Figure 3D and E). This indicates that motility is likely dispensable for survival within zebrafish and agrees with previous work showing that loss of this trait can provide bacteria with a fitness advantage in some host settings [37]. Other factors encoded within the advantageous gene set may control processes that are easily trans-complemented by neighboring bacterial cells through the production of 'common goods' or represent negative regulators of virulence determinants that promote competitive growth [38]. Thirteen of the 227 advantageous genes were hypertolerant to transposon insertion, suggesting that the inactivation of these genes may have provided a growth advantage prior to *in piscis* selection (Figure 2: advantageous).

Trait Enrichment Analysis (TEA) is used to filter Tn-seq-derived candidate genes for putative virulence determinants

To focus the list of Tn-seq-derived candidate genes for further study, we developed a qualitative ranking system using a customized algorithm referred to as **T**rait **E**nrichment **A**nalysis (TEA). The underlying premise of TEA was inspired by molecular Koch's postulates, which are a widely acknowledged set of guidelines used by investigators to critically assess the contributions made by certain genetic elements to the virulent nature of pathogenic organisms [39]. In particular, TEA draws on the first

Table 1. Biological significance of Tn-seq-derived candidate genes.

Category	Locus tag	Gene name	*p* value	$\log_2$ change	Niche	KEGG classification (subcategory)	Literature-based biological significance	Model system
Informational	*EcF11_3256*	*bipA, typA*	<0.0001	−8.12	blood,PC	Unclassified	Cold stress response [53]	*E. coli* K12 lab strain/*in vitro*
							Co-ordination of pathogenicity island gene expression [30]	EPEC/*in vitro*
							BipA binds ribosome under stress conditions [35]	*Salmonella enterica*/*in vitro*
							Stress adaptation and symbiosis [36]	*Sinorhizobium meliloti*/*in vivo* plant model
	EcF11_5238	*hdeD*	0.0128	−4.80	blood,PC	Hypothetical	Acid stress regulon [54]	*E. coli* K12 lab strain/*in vitro*
	EcF11_1255	*lrhA*	0.0001	−3.35	blood,PC	Hypothetical	Regulation of virulence genes [31]	EHEC/*in vitro*
							Regulation of motility, lipase activity, toxin expression and virulence [55]	*Xenorhabdus nematophila*/*in vivo* insect model
	EcF11_2271	*rfaH*	0.0002	−7.29	blood,PC	Informational (transcription factor)	Global regulator of virulence [32]	UPEC/*in vitro* and *in vivo* mouse model
	EcF11_4315	*nhaR*	0.0121	−1.36	blood,PC	Informational (transcription factor)	Regulates genes that facilitate cell-cell adhesion and adhesion to host cells [56]	*Proteus Mirabilis*/*in vivo* mouse model
	EcF11_0317	*sirB1*	0.0268	−3.26	blood	Hypothetical	Regulation of the SPI type III secretion system [67]	*Salmonella typhimurium*/*in vitro*
	EcF11_5302	*hdfR*	0.0019	−3.44	blood	Hypothetical	Regulator of the *std* fimbrial operon [68]	*Salmonella enterica*/*in vitro*
							Regulates flagellar synthesis [69]	*E. coli* K12 lab strain/*in vitro*
Membrane architecture	*EcF11_1390*	*kpsM*	0.0033	−8.11	blood,PC	Environmental interactions (transporter)	Capsular polysaccharide exporter [70]	*E. coli* K1/*in vitro*
							KpsM is predominantly encoded by strains of the *E. coli* B2 phylotype [71]	*E. coli*/*in silico*
							Resistance against phagocyte-mediated killing [33]	ExPEC/tissue culture
	EcF11_1396	*kpsS*	0.0229	−3.84	blood,PC	Unclassified	Capsular polysaccharide exporter [72]	*E. coli* K5/*in vitro*
	EcF11_1397	*kpsC*	0.0035	−2.57	blood	Undassified	Capsular polysaccharide exporter [73]	*E. coli* K5/*in vitro*
	EcF11_1398	*kpsU*	0.0018	−4.78	blood,PC	Metabolism (lipopolysaccharide biosynthesis)	Capsular polysaccharide exporter [73]	*E. coli* K5/*in vitro*
	EcF11_1399	*kpsD*	0.0067	−1.95	blood,PC	Hypothetical	Capsular polysaccharide exporter [74]	*E. coli* K5/*in vitro*
	EcF11_1400	*kpsE*	0.0006	−3.40	blood,PC	Metabolism (lipopolysaccharide biosynthesis)	Capsular polysaccharide exporter [57]	*E. coli*
							Capsule production and adhesion/invasion of host cells [75]	*Campylobacter jejuni*/*in vitro* and tissue culture
	EcF11_1401	*kpsF*	0.0192	−2.66	blood,PC	Metabolism (lipopolysaccharide biosynthesis)	Capsular polysaccharide exporter [76]	*E. coli* K1/*in vitro*
	EcF11_0979	*wzzB*	0.0480	−2.39	blood,PC	Metabolism (lipopolysaccharide biosynthesis)	Regulation of O-antigen chain length [77]	*Salmonella typhimurium* and *Shigella flexneri*/*in vitro*
Nutrient utilization	*EcF11_3352*	*manA*	0.0024	−3.20	blood,PC	Metabolism (fructose and mannose metabolism)	Mannose utilization, exopolysaccharide structure, biofilm formation and virulence [78]	*Photorhabdus Luminescens*/*in vitro* and *in vivo* insect model
	EcF11_2113	*fitD*	0.0048	−2.87	blood	Environmental interactions (transporter)	Facilitates iron uptake [79]	various *E. coli*/*in vitro*

Table 1. Cont.

Category	Locus tag	Gene name	*p* value	$\log_2$ change	Niche	KEGG classification (subcategory)	Literature-based biological significance	Model system
	EcF11_4915	*eutJ*	0.0191	−2.14	blood	Unclassified	Utilization of ethanolamine as a carbon source [80]	*Salmonella typhimurium*/*in vitro*
	EcF11_4920	*eutT*	0.0096	−2.49	blood	Metabolism (porphyrin and chlorophyll metabolism)	Utilization of ethanolamine as a carbon and nitrogen source [81]	*Salmonella enterica*/*in vitro*
Other	*EcF11_0498*	*tehA*	0.0188	−1.68	blood	Unclassified	Resistance to antiseptics and disinfectants [82]	*E. coli* K12 lab strain/*in vitro*
	EcF11_2210	*tieB, senB*	0.0002	−2.81	blood	Hypothetical	Enterotoxin produced by *E. coli* and *Shigella* [83]	EIEC/*ex vivo* rabbit ileum
	EcF11_3975	*ompR*	0.0164	−1.97	blood	Environmental interactions (two-component system)	Osmotic stress resistance and fitness in a mouse urinary tract infection model [84]	UPEC/*in vitro* and *in vivo* mouse model
	EcF11_2727	*potD*	0.0063	−1.44	PC	Environmental interactions (transporter)	Polyamine surface protein required for virulence [85]	*Streptococcus pneumoniae*/*in vivo* mouse model

doi:10.1371/journal.pgen.1003716.t001

axiom of these postulates, which states that, "*The phenotype or property under investigation should be associated with pathogenic members of a genus or pathogenic strains of a species.*" [40]. Thus, the primary goal of TEA is to identify correlations between bacterial genes and bacterial traits (i.e. genotype-phenotype relationships).

TEA was carried out as outlined in Figure 5. Briefly, protein sequences were acquired from an ecologically diverse, manually curated collection of 165 bacteria representing six phyla and assembled to form the TEA metaproteome database (TEA-MD, Table S3). Each bacterial strain was annotated with 'trait categories' defining habitat, niche of isolation, taxonomic lineage and phenotype (i.e. pathogen or non-pathogen). Homologues were retrieved for each of the 5,146 protein-coding genes contained within the F11 chromosome using the Basic Local Alignment Search Tool: BLASTp (Methods) [41]. Because proteins in the TEA-MD are linked to a set of traits defined by their source bacterium, homologue collections associated with F11 proteins can be appraised for the presence of a universal characteristic in a similar fashion to traditional gene ontology and KEGG enrichment analyses. As a result, hypotheses can then be generated regarding a protein's evolutionary origins (i.e. horizontal vs. vertical inheritance or homoplasy) and possible influence on bacterial physiology or behavior based on its combination of trait associations. A complete description of the TEA-MD, including the number of organisms and proteins comprising each trait category and the relative proportions of phyla compared to sequenced isolates in GenBank, is provided as supporting information (Figures S7 and S8).

Pathogen-associated genes within Tn-seq gene sets. TEA, as it was employed here, aimed to assign 'trait identities' to F11 genes, with particular attention being paid to genes that are common to known pathogens and therefore, putative virulence determinants (Figure 5). It was found that approximately 22% of the F11 genome is made up of genes with pathogenic identity—1,166 F11 genes are significantly more associated with pathogen genomes (Methods). Considering Tn-seq-derived candidate genes, it was anticipated that the *in piscis* gene set be enriched for these types of genes. Indeed, a significant portion of the 970 *in piscis* genes had pathogenic identity (246 genes, 25% of gene set, $p = 0.009$) (Table S4 and http://pathogenomics.path.utah.edu/F11_TnSeq/). The enrichment of pathogen-associated genes within the *in piscis* gene set was largely attributed to the composition of both the PC and blood gene sets, which also had a significant number of genes with pathogenic identity as determined by TEA (PC: 39 genes, ~32% of gene set, $p = 0.008$; blood: 192 genes, ~25% of gene set, $p = 0.0485$). Interestingly, the multi-niche gene set was not enriched for pathogen-associated genes (15 genes, ~20% of gene set, $p = 0.786$).

We conclude from these data that genes required for F11 fitness in multiple niches are likely to encode more basic 'core' type functions that may enable both pathogenic and non-pathogenic strains alike to better deal with general stresses encountered within varied host environments, but not necessarily in broth culture. In contrast, the niche-specific genes, which are more likely to be pathogen-associated, may endow bacteria like F11 with salient, pathoadaptive traits that are perhaps contextually regulated. Along these lines, we unexpectedly found that the advantageous gene set was also enriched for pathogen-associated genes (Table S4, 69 genes, ~30% of gene set, $p = 0.004$). We hypothesize that the disruption of these particular genes during the Tn-seq selection screen alleviated fitness costs associated with their maintenance and/or expression within the zebrafish host. Future investigation of the advantageous gene set may illuminate evolutionary tradeoffs

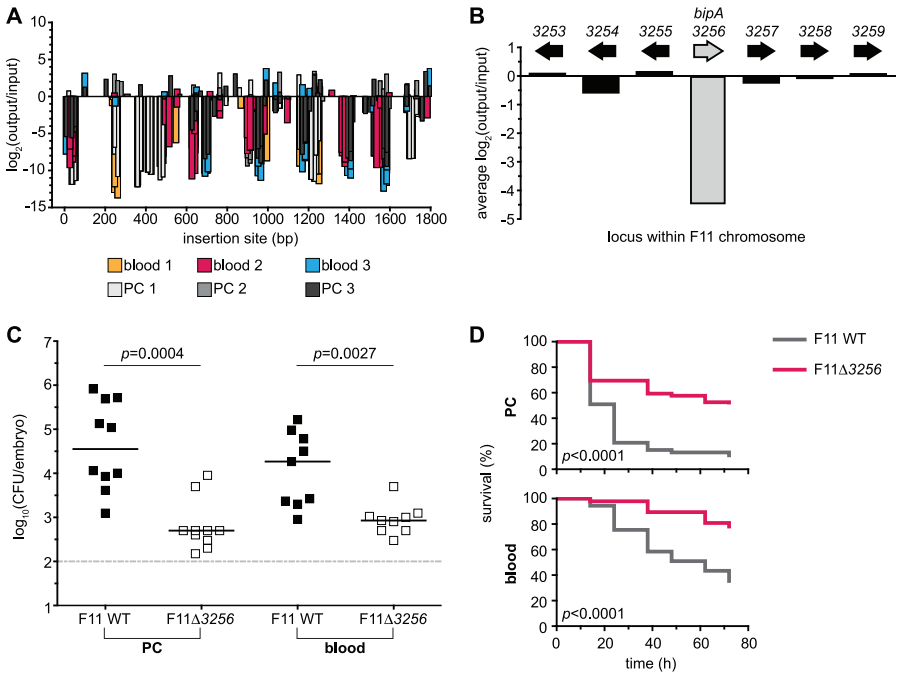


Figure 4. Retrospective deletion of the candidate gene *EcF11_3256/bipA* confirms Tn-seq as a useful tool for the identification of loci required for fitness and virulence. (A) Inserts found within the locus *EcF11_3256/bipA* are plotted with respect to their position of integration (x-axis). Magnitude and direction of bars represent the change in fitness of F11 (y-axis) that correlated with a given insertion event. Color of bars denotes pool of origin. Changes in fitness are presented as $\log_2$ of the ratio of occurrence frequencies observed between output and input pools (output/input). (B) The average of insert-based fitness changes in (A) is plotted on the x-axis (shaded bar, *bipA*). For comparison, the average alterations in occurrence frequency ratios for proximal genes are also plotted. (C) Equal numbers (1,000–2,000 CFU total) of wild type and F11Δ*3256* were inoculated into the PC (left) or bloodstream (right) of zebrafish embryos. Fish were sacrificed and bacterial loads enumerated ~20 h post-inoculation by differential plating ($n = 9$ to 10 zebrafish). Dashed line marks the limit of quantification; *p* values were determined using a paired t-test, bars indicate medians. (D) The pericardial cavity (PC, top) and blood (bottom) of 48 hpf embryos were inoculated with approximately 3,000 CFU. Fish were scored for death at 0, 12, 24, 36, 48, 60 and 72 h post-inoculation (hpi). Data are presented as Kaplan-Meier survival plots and *p* values were calculated using a log-rank (Mantel-Cox) test ($n > 45$ for each survival curve).
doi:10.1371/journal.pgen.1003716.g004

that are incurred by the acquisition and retention of these loci by pathogens as they traverse different host-associated environments.

Examination of Tn-seq candidates filtered by TEA. To investigate the effectiveness of TEA to help identify previously uncharacterized genes important for ExPEC fitness *in vivo*, we selected 3 from the subset of 246 pathogen-associated Tn-seq-derived *in piscis* genes for closer examination. *EcF11_3082*, *EcF11_2628* and *EcF11_3933* were individually deleted from the F11 chromosome and tested in competitive infection assays against wild type F11 within the zebrafish PC and blood. These genes were chosen based on their performance in the initial Tn-seq selection screen and their association with known pathogens, as determined by TEA. Trait identities for each gene are illustrated in Figure 6 along with the results from competitive infection assays.

The *EcF11_3082* gene product is homologous to the transcriptional repressor MprA/EmrR. In *E. coli* lab strains, MprA/EmrR has been shown to negatively regulate biosynthesis of the microcin B17 antibiotic and a multidrug efflux pump encoded by *emrA and emrB* [42,43]. Fifty-six strains within the TEA-MD harbor an *EcF11_3082*-like gene, 34 (61%) of which are pathogens (Figure 6A, significance of pathogenic identity: $p = 0.034$). Tn-seq indicated that *EcF11_3082* promotes the fitness of F11 in both the PC and bloodstream in all three replicates. Average fitness deficits observed through Tn-seq for *EcF11_3082* insertion mutants ranged from 5 to 21-fold. This was confirmed by one-to-one competitive assays in which survival of the *EcF11_3082* deletion mutant was markedly diminished relative to wild type F11 following inoculation of the PC and blood (Figure 6B).

Transposon insertions into the *EcF11_2628* locus decreased the fitness of F11 within the blood by an average of 15-fold, as assessed by Tn-seq. The TEA-MD contains 7 homologues of *EcF11_2628*, 6 of which belong to ExPEC strains that were isolated from a variety of host-associated niches (Figure 6C, significance of

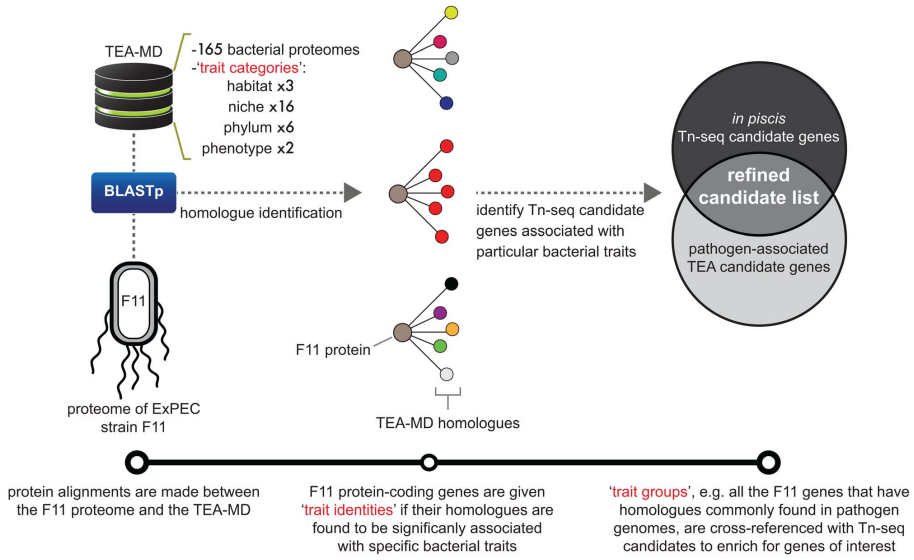


Figure 5. Diagrammatic workflow of 'Trait Enrichment Analysis' (TEA). This visualization depicts the TEA method and describes the relationships between various data types. Bottom track: linear representation of steps followed in the execution of TEA. Protein alignments are made between F11 proteins and the TEA-MD to identify homologues. Microbes within the TEA-MD are annotated with four separate traits contained within the trait categories: habitat (1 of 3), niche (1 of 16), phylum (1 of 6) and phenotype (1 of 2). The composition of traits associated with homologue sets for each F11 protein is assessed for enriched traits in order to assign trait identities. Trait groups, which are F11 genes that share a particular trait identity, are used to organize Tn-seq-derived candidate genes. TEA-MD = TEA-metaproteome database; BLASTp = Basic Local Alignment Search Tool used to align protein sequences.
doi:10.1371/journal.pgen.1003716.g005

pathogenic identity: $p = 0.035$). *EcF11_2628* encodes a hypothetical protein that is predicted by PHYRE (Protein Homology/analogY Recognition Engine) to function as a toxic extracellular metalloendopeptidase [44]. A structurally similar toxin, AsaP1, was recently described as a critical virulence factor encoded by the natural fish pathogen *Aeromonas salmonicida subsp. Achromogenes* [45]. Deletion of this gene from F11 resulted in decreased fitness within the PC and surprisingly, had no discernible effects in the bloodstream during one-to-one competition assays with the wild type strain (Figure 6D). These results run counter to those obtained with Tn-seq, where disruption of *EcF11_2628* reduced the fitness of F11 within the blood. This discrepancy may be attributable to the fact that Tn-seq is in effect a one-to-3,000 competitive assay among transposon insertion mutants within each fish, whereas in our follow-up assays specific gene knockout mutants are tested in one-to-one competition with wild type. Within these different experiments, individual mutants may be subject to very different pressures and frequency-dependent selection events.

Tn-seq results indicated that disruption of a third pathogen-associated gene, *EcF11_3933*, attenuates the fitness of F11 about 4-fold within the PC. Fifteen of the 18 EcF11_3933 homologues (83%) found in the TEA-MD are encoded by known pathogens, including multiple ExPEC strains, *E. coli* O157:H7, *Neisseria gonorrhoeae*, *Vibrio cholera*, *Haemophilus influenza*, *Photorhabdus luminescens* and *Erwinia amylovora* (Figure 6E, significance of pathogenic identity: $p = 0.001$). EcF11_3933 is uncharacterized and annotated as a homologue of the broadly conserved DNA protecting protein DprA. A related but distinct protein, DprA/Smf, has been suggested to confer competence for the uptake of foreign DNA by various bacterial species [46,47], though it is presumed to function differently in lab strains of *E. coli* [48]. In competitive assays, deletion of *EcF11_3933* rendered F11 significantly less fit than the wild type strain in both the PC and bloodstream of zebrafish (Figure 6F).

Tn-seq and TEA-derived *in piscis* candidate genes contribute to pathogen fitness within mouse models of bacteremia and urinary tract infection

To extend our experimental and sequence-based observations with the candidate genes *EcF11_3256/bipA*, *EcF11_3082*, *EcF11_2628* and *EcF11_3933*, we employed a more traditional host organism for investigating ExPEC pathogenicity—the laboratory mouse [49,50]. Evaluating Tn-seq and TEA-derived *in piscis* candidate genes using a murine model provided an opportunity to further establish the biological relevance of our multilayered approach. Murine models of bacteremia are an established means to assess the general capacity of ExPEC isolates to replicate and cause disease within a host [50,51,52]. One-to-one mixtures of wild type F11 and each mutant strain (10^8 total CFU) were subcutaneously injected into adult female Swiss Webster mice. Approximately 12 to 15 h later, bacterial titers within the spleen and liver were determined (Figure 7). Mutants lacking

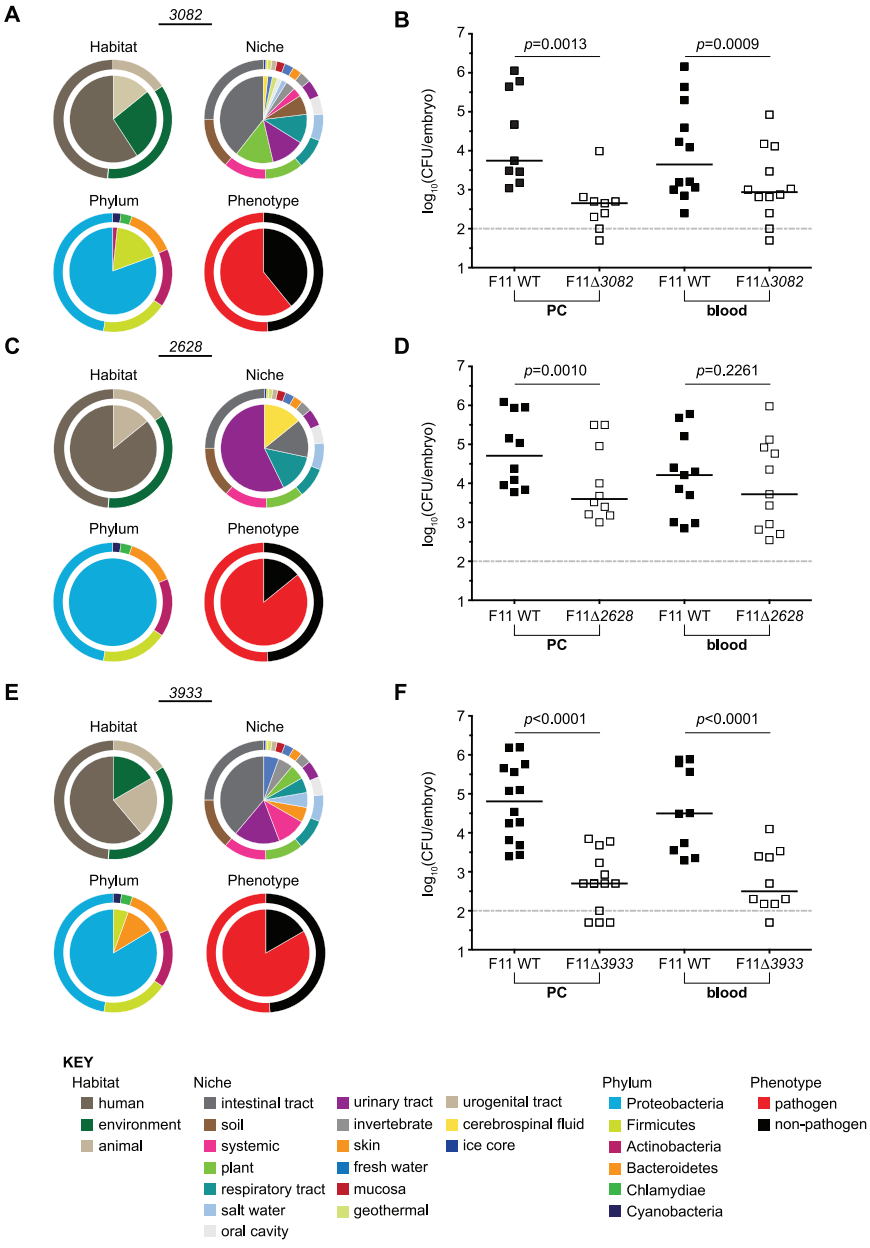

A
3082
Habitat
Niche
Phylum
Phenotype
B
p=0.0013
p=0.0009
log10(CFU/embryo)
F11 WT
F11Δ3082
F11 WT
F11Δ3082
PC
blood
C
2628
Habitat
Niche
Phylum
Phenotype
D
p=0.0010
p=0.2261
log10(CFU/embryo)
F11 WT
F11Δ2628
F11 WT
F11Δ2628
PC
blood
E
3933
Habitat
Niche
Phylum
Phenotype
F
p<0.0001
p<0.0001
log10(CFU/embryo)
F11 WT
F11Δ3933
F11 WT
F11Δ3933
PC
blood
KEY
Habitat
human
environment
animal
Niche
intestinal tract
soil
systemic
plant
respiratory tract
salt water
oral cavity
urinary tract
invertebrate
skin
fresh water
mucosa
geothermal
urogenital tract
cerebrospinal fluid
ice core
Phylum
Proteobacteria
Firmicutes
Actinobacteria
Bacteroidetes
Chlamydiae
Cyanobacteria
Phenotype
pathogen
non-pathogen

Figure 6. Targeted deletion of candidate genes with pathogenic identity attenuates the fitness of F11 within zebrafish. The composition of traits (bottom KEY) represented among bacteria encoding homologues of (A) *EcF11_3082,* (C) *EcF11_2628,* or (E) *EcF11_3933* are presented for the trait categories habitat of isolation, niche of isolation, phylum and phenotype. For reference, segmented outer rings bordering each pie graph depict the composition of traits of all genes contained within the TEA-MD. Equal numbers (1,000–2,000 CFU total) of wild type and (B) F11Δ*3082,* (D) F11Δ*2628,* or (F) F11Δ*3933* were inoculated into the PC (left) or bloodstream (right) of embryos. Fish were sacrificed and bacterial counts determined ~20 h post-inoculation by differential plating ($n = 9$ to 14 zebrafish). Dashed line marks the limit of quantification; p values were determined using a paired t-test, bars indicate medians.
doi:10.1371/journal.pgen.1003716.g006

EcF11_3256/bipA (Figure 7A), *EcF11_3082* (Figure 7B) and *EcF11_3933* (Figure 7D) were significantly outcompeted by the wild type F11 parent strain. However, F11Δ*3256* exhibited a more modest defect in the spleen ($p = 0.0828$). In contrast, there was no observable competitive difference between wild type and F11Δ*2628* in either organ (Figure 7C).

Of note, we found that F11Δ*3256* displayed a cold sensitivity phenotype when grown at 20°C in broth culture, which is consistent with reports on BipA function in other bacteria (Figure S9A) [53]. Therefore, we were initially concerned that the growth rate of this mutant at 28.5°C—the temperature used to raise zebrafish embryos—was compromising our ability to distinguish its temperature and host-dependent phenotypes. However, F11Δ*3256*, like the other mutants, grew similar to the wild type strain at both at 28.5°C and 37°C (Figure S9). We conclude that the attenuated fitness phenotypes observed with F11Δ*3256* in competitive assays in both zebrafish and mice are independent of temperature effects.

We next tested the fitness contributions made by each of the four candidate genes above within the murine urinary tract, which is another commonly used model system for understanding ExPEC pathogenicity. Three days after inoculation of one-to-one mixtures of wild type and mutant derivatives (10^7 total CFU) into the urinary tract of adult female CBA/J mice, kidneys and bladders were harvested and bacterial titers determined. Of the four genes assayed, only *EcF11_3933* was critical for the competitive fitness of F11 within the bladder and kidneys (Figure 7E).

Our ability to identify candidate genes necessary for ExPEC fitness within mammalian host niches through initial prescreening in embryonic zebrafish highlight the utility of this teleost host to model overlapping aspects of ExPEC biology. By comparing the specific host and niche-specific phenotypes of various ExPEC isolates and mutant strains in futures investigations, we will be able to gain a deeper understanding of the host-pathogen interactions engaged by ExPEC.

Discussion

Application of Tn-seq to investigate the multigenic basis of ExPEC pathogenicity

The genetic and phenotypic plasticity of ExPEC isolates, which is in part due to horizontal transfer of pathoadpative accessory genes, presumably enables these pathogens to traverse a wide range of environmental and host-associated niches [7,8,10,11]. It will be important to decipher how ExPEC employ their genetic repertoires so that more specific and efficacious therapeutic strategies can be developed. Moreover, knowledge of key events that occur during the evolutionary assembly of pathogen genomes will likely be gleaned as genotype-phenotype relationships are mapped. To this end, we functionally dissected the contents of an ExPEC reference genome using the recently described transposon-based screening technique Tn-seq (Figure 1) [14,15]. With mutant pools derived from the ExPEC reference isolate F11 and multi-niche zebrafish infection models, we found that ~18% of the F11 genome significantly affected bacterial fitness under the patho-genic conditions tested. These fitness determinants can be grouped based on the niches where they were required, which in some instances correlated with particular functions like iron transport and amino acid metabolism (Figure 3). Gene set and gene-based contributions to fitness can be explored through an online and curated data viewer (http://pathogenomics.path.utah.edu/F11_TnSeq/). Going forward, it is our intent that this work will expedite future comparative and functional pathogenomic analyses focused on elucidating the genetic basis of pathogen behavior and evolution.

Seventy-six genes (1.4% of genome content) were found to significantly influence the fitness of F11 in both the PC and blood of zebrafish. We suggest that these multi-niche genes largely provide core functions that are critical, but not necessarily specific, to pathogen physiology. It was not surprising that several of these genes were recognizable and annotated as being involved in the regulation of gene expression and membrane architecture—two functions that are important to pathogens and non-pathogens alike (Table 1). In addition to the translational regulator EcF11_3256/BipA, which we validated in follow up experiments using a targeted deletion mutant and multiple *in vivo* host model systems (Figure 4 and 7A), several other informational genes, including *hdeD*, *lrhA*, *rfaH* and *nhaR*, were also identified by our Tn-seq screens as being important to the fitness of F11 within multiple niches. These genes have been shown to affect acid stress resistance, toxin expression, adhesion to host tissues and virulence gene regulation in a diverse array of pathogens, including enterohemorrhagic *E. coli* (O157:H7), *Xenorhabdus nematophila*, *Proteus mirabilis* and a sub-group of ExPEC isolates like F11 known as uropathogenic *E. coli* (UPEC) (Table 1) [31,32,54,55,56]. Additionally, genes responsible for the production of capsule—which is a well-established virulence feature that serves as a protective polysaccharide barrier surrounding the bacterial cell wall—were also important in both the PC and blood (Table 1) [57]. Further identification and, importantly, functional characterization of multi-niche genes will lead to a deeper understanding of the general fitness and virulence gene networks employed by ExPEC and other virulent bacteria within varying host environments.

Aside from multi-niche genes, we observed a striking numerical imbalance of genes that are exclusively required for either blood-borne or tissue-localized PC fitness (772 and 122 genes, respectively). The large number of genes and variety of enriched KEGG functional categories within the blood-specific gene set corroborated previous experimental observations that indicated the blood of embryonic zebrafish is a hostile environment and highly restrictive to bacterial growth [10,11]. Specifically, genes involved in nutrient acquisition and utilization made up a significant portion of the blood-specific genes (Figure 3). These results provide evidence that during systemic infections nutrients are more limiting than during localized infection of the PC, requiring ExPEC to employ a greater portion of their genetic repertoire to maintain fitness in the blood. However, additional factors could also account for the discrepancy between the PC and blood-specific gene sets, including instances of trans-complementation. Once injected into the PC, bacteria remain confined until

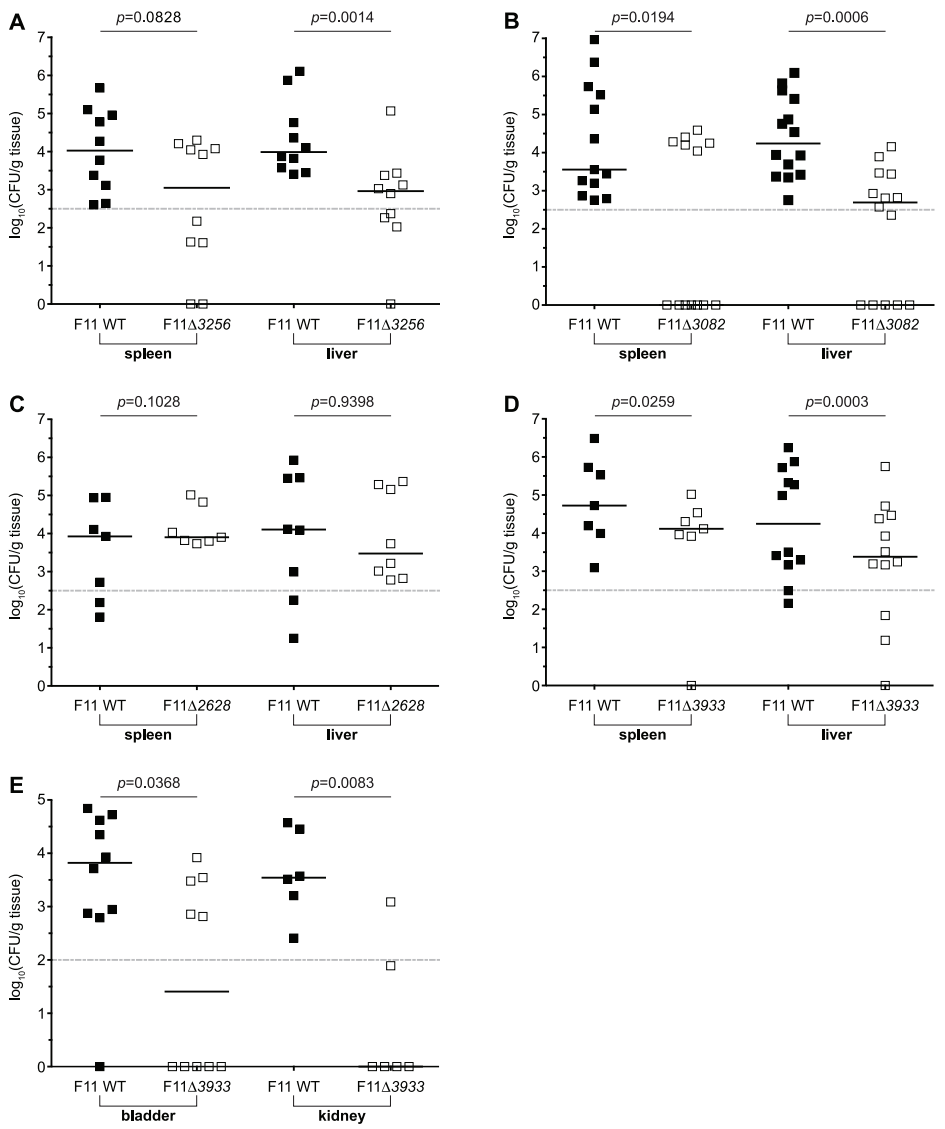


Figure 7. Tn-seq and TEA-derived candidate genes promote the fitness of F11 within murine models of bacteremia and urinary tract infection. Equal numbers (10^8 total CFU) of wild type F11 and (A) F11Δ*3256*, (B) F11Δ*3082*, (C) F11Δ*2628*, or (D) F11Δ*3933*, were subcutaneously injected into the nape of the neck of adult female Swiss Webster mice. Approximately 12 to 15 h later, bacterial titers within the spleen (left) and liver (right) were determined by differential plating. (E) Equal numbers (10^7 total CFU) of wild type F11 and F11Δ*3933* were transurethrally inoculated into the urinary tract of adult female CBA/J mice. Bacterial titers present in the bladder (left) and kidneys (right) were determined 3 d post-inoculation by differential plating. Dashed line marks the limit of quantification. $n = 7$ to 14 mice; p values were determined using a paired t-test, bars indicate medians.
doi:10.1371/journal.pgen.1003716.g007

the embryo succumbs to the infection [10]. Because bacteria stay in close proximity to one another within the PC, the chances for trans-complementation and the sharing of common goods are high and may result in relaxed selection. A recent study demonstrated that extensive trans-complementation can occur within mixed mutant populations of *Yersinia pestis* in a mouse model of pneumonic plague [58]. However, the authors noted that this appears to be a unique characteristic of *Y. pestis*, as other pathogens with tropism for lung tissue do not exhibit such a high degree of wild type-to-mutant rescue. In our assays, trans-complementation effects within the PC were apparently limited as we were able to identify genes required for ExPEC survival within this niche.

Ultimately, this initial genome-wide analysis of the components necessary for ExPEC fitness within pathogenic environments garnered two key insights: (i) the number of genes that influence fitness in multiple niches is relatively small and (ii) there are distinct subsets of genes employed by ExPEC within different host environments. This suggests that *do-it-all* genes do not necessarily dictate the generalist and plastic nature of ExPEC. Instead, ExPEC may evolutionarily maintain sets of context-specific genes that can be employed, possibly in different combinations, to respond in a highly specific-manner to a variety of conditions. This characteristic likely contributes to the seemingly redundant composition and versatility of contemporary ExPEC genomes. Further functional dissection of ExPEC genomes by Tn-seq-based approaches will help clarify which genes underlie the salient traits of this lineage and inform the design of broad-spectrum therapies that target only ExPEC while leaving beneficial or benign *E. coli* strains within the normal microbiota unperturbed.

Refinement of candidate gene sets by TEA

A longstanding challenge associated with unbiased, high-throughput functional genomics techniques like Tn-seq has been the need to develop methods that provide an effective transition to focused follow-up studies. Much work has been done to quantitatively filter datasets using sophisticated determinations of statistical significance and false discovery control procedures. Analysis of the molecular pathways represented within functional genomics gene sets has also been addressed extensively. However, less attention has been given to systematic methods that exploit the qualitative features of individual genes and how such considerations can streamline candidate gene vetting. To address this issue, we developed a **T**rait **E**nrichment **A**nalysis (TEA) tool to organize Tn-seq-derived candidate genes in a biologically meaningful way (Figure 5). With a manually curated set of 165 microbial proteomes, TEA is able to assign meta-features to F11 genes based on their association with specific bacterial traits (i.e. habitat, niche of isolation, phylum and phenotype). In this way, TEA provides a tractable method to categorize Tn-seq candidate genes into a more focused subset—something not easily accomplished using larger microbial sequence databases like GenBank and BioCyc or functional annotation platforms such as DAVID. TEA uncovered several genes with previously unappreciated roles in ExPEC pathogenicity, thus creating new opportunities to further understand the genetic underpinnings of this lineage.

For the purposes of our investigation, we used TEA to identify F11 genes with homologues that are more often associated with the genomes of known pathogens. We describe such genes as having 'pathogenic identity' and reason that, because of their affiliation with pathogens, they are more likely to contribute to pathogenic behavior—a principal tenet of molecular Koch's postulates [39,40]. In considering the overlap between Tn-seq-derived *in piscis* genes and the TEA-derived gene set made up of F11 genes with pathogenic identity, we were able to identify 246 genes that potentially influence the *in vivo* fitness capacity of F11 specifically. We selected three of these pathogen-associated candidate genes for closer inspection.

The candidate gene *EcF11_3082* was identified by Tn-seq and confirmed during one-to-one co-challenge experiments to be important for the fitness of F11 within the blood and PC of zebrafish and spleen and liver of mice (Figure 6B and 7B). This gene appears to be relatively widespread within the TEA-MD, being found in 56 genomes, 61% of which are classified as pathogens. The range of bacteria encoding homologues of EcF11_3082 includes isolates from 12 (out of 16) different niches and 3 (out of 6) phyla represented in the TEA-MD (Figure 6A). These observations suggest that *EcF11_3082* may play a broad role in bacterial fitness within a variety of pathogenic and non-pathogenic contexts. *EcF11_3082* is predicted to encode an MprA/EmrR-like transcriptional repressor, which was shown to regulate the expression of a downstream multidrug efflux pump encoded by *emrA* and *emrB*. Although our Tn-seq selection screen and co-challenge experiments identified EcF11_3082 as a fitness determinant, disruption of the *emrA* and *emrB* genes present in F11 did not appreciably affect fitness within the zebrafish host, as determined by Tn-seq. This suggests that either disruption of *EcF11_3082* led to aberrant and deleterious expression of *emrAB* that resulted in attenuation of the *EcF11_3082* mutant, or the particular MprA/EmrR allele harbored by F11 influences other aspects of bacterial physiology not yet defined. Further investigation is required to determine the exact influence of *EcF11_3082* on pathogenicity.

EcF11_2628, which encodes a 'conserved hypothetical protein' that has some homology to a secreted metalloendopeptidase toxin, was also identified by Tn-seq and confirmed as an important fitness determinant using a targeted knockout strain (Figure 6D). However, we were unable to ascertain if this gene is required during colonization of a mammalian host. The F11Δ*2628* mutant variant did not exhibit any statistically significant decline in competitive fitness during systemic or urinary tract infection of mice (Figure 7C and data not shown). Further investigation is required to determine the physiological context in which this putative toxin promotes ExPEC fitness. Nonetheless, homologues of *EcF11_2628* were found in only 7 proteobacterial genomes within the TEA-MD (Figure 6C and D). Of these, 5 are ExPEC strains that were isolated from human patients with either meningitis (strain S88) or UTI (strains CFT073, UTI89, 536 and UMN026). Another ExPEC strain (APEC01) harboring an *EcF11_2628* homologue came from a case of avian sepsis that occurred subsequent to respiratory tract infection. The sole non-pathogen that carried an *EcF11_2628* homologue was ED1a, a commensal isolate from the human gut that is classified as belonging to the B2 phylogenetic group—the *E. coli* lineage that encompasses the vast majority of human ExPEC strains [7]. This observation is intriguing because the *EcF11_2628* allele carried by ED1a could represent either a genetic vestige from a pathogenic past or a new acquisition that is edging ED1a down an evolutionary path towards a more ExPEC-like countenance. Elucidating the function of this hypothetical gene should help explain its seemingly restricted distribution within a phylogenetically narrow range of bacteria.

The final pathogen-associated gene that we examined, *EcF11_3933*, encodes a variant of the DNA protection protein DprA [46,47]. However, this particular *dprA* allele is distinct from those previously characterized in other bacteria. Within the TEA-MD, 83% of the *EcF11_3933* homologues were found in the genomes of known pathogens. Interestingly, pathogens harboring an *EcF11_3933*-like gene included ExPEC isolates as well as other members of Proteobacteria and species within the phyla Bacter-

oidetes and Firmicutes (Figure 6E). In our Tn-seq selection screens, *EcF11_3933* was initially identified because it appeared to affect fitness within the PC, but subsequent analyses revealed that *EcF11_3933* could also promote pathogen fitness within the bloodstream of both zebrafish and mice (Figure 6F and 7D). Indeed, there were indications that *EcF11_3933* contributed to blood-borne fitness, but it was filtered out of the final Tn-seq-derived blood gene set because it did not meet all of the predetermined cutoffs, though only by a thin margin. These results identify limits to Tn-seq as it is applied here, but also highlight the value of using multiple infection models to discover and assess the roles of fitness and virulence determinants.

Additional work using a murine UTI infection model demonstrated that *EcF11_3933* is also important for competitive fitness within the murine bladder and kidneys (Figure 7E). This is in contrast to *EcF11_3256*, *EcF11_3082* and *EcF11_2628*, which we found to be dispensable for colonization of the urinary tract (data not shown). Interestingly, alleles of *c4222*, which is a homologue *of EcF11_3933* harbored by the UPEC strain CFT073, were shown by microarray to be transcriptionally induced within a panel of UPEC isolates during active infection of the human urinary tract [59]. Together, previous observations and those presented here indicate that *EcF11_3933* is of broad importance to the pathogenic potential of F11 and other ExPEC isolates within a variety host-associated niches. Moreover, the distribution of *EcF11_3933*-like genes among a diverse array of pathogens (e.g. ExPEC, the enteric pathogen *E. coli* O157:H7, *Neisseria gonorrhoeae*, *Vibrio cholera*, *Haemophilus influenza*, *Photorhabdus luminescens* and *Erwinia amylovora*), which can colonize and cause disease within a broad range of animal, insect and plant hosts, suggests that *EcF11_3933* also represents a multi-lineage virulence determinant. Further analysis of *EcF11_3933* and its pathogen-associated homologues may prove useful in defining common virulence strategies employed by ExPEC as well as other more distantly related bacterial species.

Concluding remarks

Cumulatively, this investigation demonstrates that the application of Tn-seq, coupled with TEA, provides an efficient means to identify novel or previously unappreciated fitness and virulence determinants. The utility of TEA could easily be enhanced through expansion of the TEA-MD by, for example, the addition of metadata obtained from other host-microbe experimental systems. TEA serves as an accelerated annotation layer for bridging high-throughput sequencing experiments and functional characterization. In moving forward, it will be important to continue the careful curation of pertinent biological information associated with the rapidly expanding and often redundant and ill-defined collections of microbial genomics data so that meaningful connections can be mined. Ultimately, the use of Tn-seq and TEA with an expanded number of bacterial isolates and host systems may provide a detailed map of the gene families and signaling cascades employed by ExPEC and other pathogens to colonize host-associated environments.

Methods

Ethics statement

Animals used in this study were handled in accordance with University of Utah approved IACUC protocols that followed the standard guidelines described at www.zfin.org and in the Guide for the Care and Use of Laboratory Animals, 8th Edition.

Bacterial strains and plasmids

All bacterial strains and plasmids used in this study are listed in Table S5. Unless specified otherwise, bacteria were cultured statically at 37°C for 24 h in 20 ml of a defined M9 minimal medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1% glucose, 0.0025% nicotinic acid, 0.2% casein amino acids and 16.5 mg/ml thiamine in H_2O) prior to injection into zebrafish embryos or mice. Antibiotics (kanamycin or ampicillin) were added to the growth medium when necessary to maintain recombinant plasmids or select for mutants.

Targeted gene knockouts of Tn-seq-derived candidate genes were generated in ExPEC strain F11 using the lambda Red-mediated linear transformation system [60,61]. Briefly, a kanamycin resistance gene was amplified using polymerase chain reaction (PCR) from the pKD4 plasmid with 40-base pair overhangs specific to the 5′ and 3′ ends of each targeted locus. PCR products were introduced via electroporation into F11 carrying pKM208, which encodes an IPTG (isopropyl-β-D-thiogalactopyranoside)-inducible lambda red recombinase. Knockouts were confirmed by PCR. Primer sets used are listed in Table S6.

Retrofitting of the previously constructed pSAM_Bt vector to be used for the transposon mutagenesis of ExPEC isolate F11 was accomplished using source vectors and primers as listed in Tables S5 and S6, respectively. The P_{tac} promoter from pGFP-Mut3.1 (Clonetech), along with its associated ribosome binding sequence, was amplified via PCR. Engineered 5′ BamHI and 3′ NdeI restriction sites were used to sub-clone the resulting fragment into a BamHI/NdeI (New England Biolabs) double digested pSAM_Bt vector upstream of the *himar*1C9 transposase gene. The kanamycin resistance gene from pKD4 was amplified and ligated using the restriction sites MfeI and XbaI, replacing the erythromycin resistance gene *ermG* in pSAM_Bt. The resulting transposon mutagenesis vector, pSAM-Ec, was stored and propagated in the *pir*$^+$ *E. coli* strain EcS17.

Zebrafish embryos

*AB wild-type zebrafish embryos were collected from a laboratory-breeding colony that was maintained on a 14 h/10 h light/dark cycle. Embryos were grown at 28.5°C in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.4 mM $CaCl_2$, 0.16 mM $MgSO_4$) containing 0.000016% methylene blue as an anti-fungal agent.

Infection of zebrafish

One ml from a 24-h bacterial culture of each transposon mutant library or isogenic bacterial strain was pelleted, washed once with 1 ml sterile PBS (Hyclone) and re-suspended in approximately 1 ml PBS to obtain appropriate bacterial densities for microinjection. Prior to injection, 48 h post-fertilization (hpf) embryos were manually dechorionated, briefly anesthetized using 0.77 mM ethyl 3-aminobenzoate methanesulfonate salt (tricaine) (Sigma-Aldrich) and embedded in 0.8% low-melt agarose (MO BIO Laboratories) without tricaine. Approximately 1 nl of bacteria was injected directly into the pericardial cavity (PC) or the blood via the circulation valley (CV) located ventral to the yolk sac using a YOU-1 micromanipulator (Narishige), a Narishige IM-200 microinjector and a JUN-AIR model 3-compressor setup. For each experiment, average colony forming units (CFU) introduced per injection were determined by adding 10 nl of the inoculum into 1 ml 0.7% NaCl that was then serially diluted and plated on Luria-Bertani (LB) agar plates. For co-challenge experiments, input doses were plated on LB agar +/− kanamycin (50 μg/mL) to determine relative numbers of the wild type and mutant strains present. After injection, embryos were carefully extracted from agar and placed in either a 10 cm petri dish containing E3 medium for transposon screening or, individually into wells of a

96-well microtiter plate (Nunc) containing E3 medium for co-challenge experiments. The E3 medium used during this incubation period lacked both tricaine and methylene blue.

To harvest DNA from embryos infected with transposon mutant libraries at the end of the selection period, all infected fish were collected (both alive and dead) in a 1.6 ml tube, being careful to remove as much E3 medium as possible by gently pulsing the tube in a microfuge to facilitate separation. Embryos were suspended in 500 µl of 0.5% Triton X-100 in PBS and homogenized using a mechanical PRO 250 homogenizer (PRO Scientific). Homogenates were then centrifuged at 18,000 g for 5 minutes to sediment bacteria away from host cellular components released during Triton X-100-mediated lysis. The resulting pellet, which appears black because of remaining zebrafish debris, served as the starting material for DNA extraction using the Wizard Genomic DNA Purification Kit (Promega) as per the manufacturer's protocol with a brief modification: the duration of the initial lysis step was extended as needed to dissolve the tough pellet.

To quantify bacterial numbers at the completion of co-challenge experiments, embryos were homogenized at the indicated time points in 500 µL PBS containing 0.5% Triton X-100 using a mechanical PRO 250 homogenizer. Homogenates were then serially diluted and plated on LB agar +/− kanamycin (50 µg/mL) to determine relative numbers of wild type and mutant bacteria. For co-challenge assays, only embryos where the total bacterial CFU recovered was equal to or greater than the limit of quantification (i.e. at least 20 CFU counted on the lowest serial dilution, which corresponds to 100 CFU per embryo) are reported. For lethality assays, fish were inspected for death at the indicated time points over a 72 h period. Death is defined here as the complete absence of heart rhythm and blood flow. Survival graphs depict total pooled results from at least 3 independent experiments in which groups of approximately 15 to 20 embryos were injected at a time.

Mouse infections

For bacteremic co-challenge experiments, seven to eight-week old female Swiss Webster mice (Charles River) were anesthetized using isoflurane inhalation and injected subcutaneously via the nape of the neck with 200 µl of a 1:1 wild type to mutant bacterial suspension containing a total of 10^8 bacteria in sterile PBS. Spleens and livers recovered 12 to 15 h later were weighed and homogenized in 500 µl sterile PBS. Homogenization was done with a Bullet Blender Storm 24 (Next Advance) using three SSB32 3.2 mm stainless steel beads for 5 minutes on power level 6. Homogenates were serially diluted and plated on LB agar +/− kanamycin (50 µg/mL) to determine the number of both wild type and mutant bacteria. Bacteremic co-challenge experiments were repeated at least twice. For these assays, only organs where the total bacterial CFU recovered was equal to or greater than the limit of quantification (i.e. at least 20 CFU counted on the lowest serial dilution, which corresponds to approximately 200 to 600 CFU per g organ) are reported.

For co-challenge experiments within the murine urinary tract infection, seven- to nine-week old female CBA/J mice (Jackson Labs) mice were anesthetized using isoflurane inhalation and inoculated via transurethral catheterization with 50 µl of a 1:1 wild type to mutant bacterial suspension containing a total of 10^7 bacteria suspended in PBS. Bladders and kidneys were recovered 3 days later and each was weighed and homogenized in 1 ml containing 0.025% Triton X-100. Homogenates were serially diluted and plated on LB agar +/− kanamycin (50 µg/mL) to determine number of both wild type and mutant bacteria.

Statistical analysis of zebrafish and mouse infections

For co-challenge experiments, numbers of wild type and mutant bacteria present in the inoculum and recovered from host tissues were enumerated by differential plating on selective media as described above. To determine if bacterial counts differed between wild type and mutant-infected animals, a paired t-test was performed on $\log_{10}$-transformed values. Graphs and statistics were generated using GraphPad Prism 5.

Transposon library generation

Transposon mutagenesis. A detailed protocol for the transposon mutagenesis of ExPEC strain F11 using the *E. coli* donor strain EcS17 carrying pSAM-Ec and the methods to verify and sequence mutant libraries is provided in Protocol S1. Briefly, EcS17/pSAM-Ec and F11 were mixed 2:1 (donor:recipient), deposited onto nitrocellulose 0.45 µm filter discs (Millipore) and incubated for 5 h on agar plates at 37°C. Because the P_{tac} promoter allows for constitutive low-level expression of the *himar*1C9 transposase, no induction was required. Bacteria from individual mating discs were recovered in 2 ml of 1× M9 salts by vortexing. A 100 µl aliquot was serially diluted and plated on selective agar plates to determine mutagenesis frequency. The remaining 1.9 ml of mating mixture was added to 20 ml of selective media and allowed to grow shaking at 37°C until an optical density at 600 nm (OD_{600}) of ~0.5 was reached. One ml aliquots from individual matings were then store at −80°C until used in selection screening. In the event a mutant library did not contain a satisfactory number of mutants, for example, <50,000 mutants, frozen axillary stocks were briefly thawed, combined and refrozen.

Bioinformatic analysis

Sequence alignment. Sequence reads were trimmed to remove transposon sequence and the remaining 12 bp sequences were aligned to the F11 genome consisting of 119 contigs (GenBank: AAJU00000000). The alignment was preformed with Bowtie [62], allowing no mismatches to the trimmed 12-mers. The number of times each 12-mer was observed (abundance) and the insertion position were tracked in a relational database. Because our strategy for determining candidate regions relied on the Wilcoxon signed-rank test, which is relatively resistant to spurious observations, there was no limit to how many times a given 12-mer was allowed to align to the genome.

Candidate identification. To distinguish loci within the F11 genome that were hypo or hyper-tolerant to transposon insertion, the average number of inserts observed within each annotated region, considering input pools 1, 2 and 3 alone, was calculated. A Z-test was performed on the population of 5,312 loci—those that fell below a standard score of −1 were considered hypo-tolerant and those that were above a standard score of +1 were considered hyper-tolerant. Standard scores were used here as an intuitive demarcation within the distribution of insertion tolerances and are not necessarily used to indicate any particular statistical significance. Of note, because we allowed reads to be aligned to the genome multiple times, we acknowledge that the specific values calculated for insertion tolerances for each annotated F11 region is overestimated. Indeed, we found that on average, reads could be aligned approximately twice to the F11 chromosome (distribution median = 2). This accounts for the elevated insertion rates reported in Figure 2 and Table S2, which are approximately two times greater than the inferred genome-wide insertion rate of 1 in 75 bp that is reported in the results section. Although this approach added a source of noise, it also allowed us to quantify and rank regions that would have otherwise been assigned a '0'

(i.e. putative essential genes). The design of future Tn-seq experiments should evaluate the pros and cons of such an approach.

To calculate changes in fitness *in vivo* due to transposon insertion, occurrence frequencies for insertion variants were first determined by normalizing insert abundance. The normalized occurrence rate of each insertion was obtained by dividing the number of times an insert was observed by the sum of insert counts for the insert positions that were detected in both the input and output pools of the replicate screen in question. Next, a Wilcoxon signed-rank test was performed on each F11 locus to compare the distribution of occurrence frequencies of inserts found within both input and respective output pools. Because this basic approach did not immediately account for inserts that were absent from output pools, which possibly indicates a strong negative selection event, we assigned the limit of detection (1/sequencing depth) to *lost* inserts that were present within a cognate input pool and within the top 1% of occurrence frequencies of all inserts that were *lost*. By incorporating these *lost* inserts prior to performing the Wilcoxon signed-rank test, we could more confidently account for inserts that likely represented a true negative selection signal while not considering inserts that were absent due to sampling error or passive loss. Regions having significantly altered frequency distributions ($p<0.05$) and at least 10 measured inserts were assigned to the *in piscis* gene set if the average occurrence frequency of inserts within the output samples was reduced by at least 2-fold, or assigned to the advantageous gene set if the average occurrence frequency of inserts within the output samples was greater than that of the input by at least 2-fold. Using these orthogonal vetting parameters we calculated that the false discovery rate (FDR), as defined by Benjamini and Hochberg, 1995, for the collection of observations with a maximal p value of 0.05 to be 0.096 and 0.094 for the *in piscis* and advantageous gene sets, respectively [63]. Figure S5 provides a diagrammatic summary of the bioinformatic pipeline used above to determine Tn-seq candidate genes and gene sets. Table S7, along with an online data viewer (http://pathogenomics.path.utah.edu/F11_TnSeq/), can be used to anonymously access and browse final metric-based data that were generated using these methods. Data analyses were done using custom software written in Python and the open source math tool SciPy.

KEGG-based functional analysis of gene sets. Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologies (KO) were assigned to genes within the F11 genome using the KEGG automatic annotation server (KAAS) [64,65,66]. Genes without KO assignments were considered to have a 'hypothetical' function and genes encoding RNAs (e.g. tRNAs, non-coding RNAs and ribosomal RNAs) were manually assembled into a custom 'RNA' orthology. To test for enrichment of a particular KEGG functional category within a gene set, Monte Carlo simulations were performed by sampling an equivalent number of genes from the F11 genome at random 10,000 times without replacement. The functional composition of observed gene sets was then compared to theoretical sample sets using a Z-test. If a particular functional category was significantly overrepresented compared to chance ($p<0.05$) it was considered enriched. Data analyses were done using custom software written in Python and the open source math tool SciPy.

Trait enrichment analysis (TEA). 165 bacterial proteomes were downloaded from the National Center for Biotechnology Information's (NCBI) microbial genome sequence collection and assembled into the TEA metaproteome database (TEA-MD) (Table S3). Each strain included in the TEA-MD was annotated for habitat, niche of isolation, phylum and phenotype. Annotations were retrieved manually by referring to original sequencing project resources. Homologues within the TEA-MD were identified for each protein-coding gene within the F11 genome using the Basic Local Alignment Search Tool BLASTp [41]. Homologue assignment was conferred if bi-directional alignments produced an E value$<10^{-6}$ with similarity over 50% of query length. The resulting set of TEA-MD microbes harboring homologues of specific F11 proteins were then used to determine particular trait associations. For each F11 protein, 1,000 Monte Carlo simulations were performed by sampling an equivalent number of bacterial genomes from the TEA-MD without replacement. Genes that were found to have significant trait enrichments ($p<0.05$) were also required to have at least 5 TEA-MD homologues, 60% of which belonged to pathogen genomes, to be further considered and placed into trait groups. TEA was performed in parallel and independent of Tn-seq. Cross-referencing of several well-known bacterial virulence factors, in addition to broadly conserved genes involved in basic bacterial physiology, confirmed that our TEA approach could reliably distinguish pathogenic identity between these two gene classes (Table S8). Data analyses were done using custom software written in Python and the open source math tool SciPy.

Supporting Information

Figure S1 Plasmids used in this study. A) pSAM_Bt was originally constructed to express the *himar1C9* transposase (light green open reading frame) within *Bacteroides thetaiotaomicron*, facilitating insertion of the *ermG* erythromycin resistance gene (blue open reading frame) into the chromosome this bacterium. (B) pSAM_Bt was retrofitted using the indicated restriction sites with a kanamycin resistance gene (purple open reading frame) and a P_{tac} promoter to drive expression of the *himar1C9* transposase within *E. coli*. This plasmid, pSAM-Ec, was then transferred into *E. coli* strain F11 via conjugation. Other pSAM-Ec features include: two transcriptional terminators downstream of the kanamycin resistance gene (red stop signs), P7 priming sites for Illumina sequencing (dark green blocks), MmeI-modified restriction sites (red blocks) flanking the transposon (grey), a β-lactamase gene for donor selection (red open reading frame) and an RP4 oriT/oriR6K origin of replication (orange block) for specific propagation in *pir*-positive donor strains.
(PDF)

Figure S2 Assessment of transposon mutant pool diversity and saturation. (A) *E. coli* F11 transposon mutants were randomly isolated and subjected to Southern blot analysis. Genomic DNA was isolated, digested with the restriction enzyme HindIII, resolved by gel electrophoresis and probed with a digoxigenin-labeled probe specific for the kanamycin resistance gene within the transposon. (B to D) Representative colorimetric microbiological agar plates were used to estimate mutant occurrence frequency within mutant pools. (B) *E. coli* F11 transposon mutants grown on MacConkey agar to assay for the presence of mutants deficient for lactose utilization (white colony within inset). (C) Agar plates containing the dye Congo red were used to determine the frequency of insertion variants that disrupted normal curli production. Inset shows a white colony (not able to bind Congo red due to the presumed absence of curli) surrounded by curli producing mutant variants (red colonies). (D) Mutant colonies grown on Kornberg agar and subsequently exposed to iodine vapor to stain for the presence of glycogen. Normal colonies exhibit a brown coloration, colonies lacking glycogen synthesis are yellow-white (left inset) and colonies overloaded with glycogen are dark brown or black (right inset).
(PDF)

Figure S3 Diagram of a 48 h post-fertilization zebrafish embryo. *E. coli* F11 transposon mutants were delivered into zebrafish embryos via one of two injections sites. The pericardial cavity (PC) simulates a localized infection—bacteria are restricted to the indicated area (green). The circulation valley (CV) is used to deliver bacteria systemically throughout the bloodstream (red).
(PDF)

Figure S4 Virulence of *E. coli* F11 within zebrafish embryos. (A) Approximately 2,000 colony-forming units of *E. coli* F11 were delivered into the pericardial cavity (PC, left) or the blood (right). Survival of embryos was monitored over a three-day period ($n = 20$ embryos for each curve). (B) At 21 h post-inoculation, infected zebrafish were homogenized and bacterial titers determined. Each square symbol represents an individual zebrafish and bars mark median values ($n = 10$ embryos).
(PDF)

Figure S5 Description of pipeline used to define Tn-seq candidate gene sets. Diagram describes each of the five steps in determining Tn-seq-derived candidate genes and gene sets. (I) normalize insert read counts, (II) identify strong negative selection events and median-center output data, (III) perform Wilcoxon signed-rank test, (IV) initial candidate list vetting and (V) final gene set construction and curation.
(EPS)

Figure S6 Relative contributions made by each replicate screen to the *in piscis* gene set. For clarity, the *in piscis* gene set was divided into the component genes sets (A) blood, (B) PC and (C and D) multi-niche. The proportion of genes identified from only one replicate screen (numbered 1, 2 and 3) is represented by dark wedges, whereas colored wedges indicate genes that were identified from two or more screens using Tn-seq (bottom key). Fischer's exact test was used to determine if there was significant overlap between genes contributed by each replicate screen within each gene set (p values indicated where significance was found).
(PDF)

Figure S7 Compositional features of the TEA metaproteome database (TEA-MD). Pie graphs depict the proportion of protein sequences contained within the TEA-MD that are contributed by bacteria annotated with the indicated traits under the trait categories (A) habitat of isolation, (B) niche of isolation, (C) phylum and (D) phenotype. (E) Table detailing the total number of bacterial genomes and respective aggregate of protein sequences represented by specific traits within each trait category.
(PDF)

Figure S8 Comparison of phylum representation between NCBI and the TEA-MD. The number of genome sequences currently reported in the National Center for Biotechnology Information (NCBI) database for Proteobacteria, Bacteroidetes, Firmicutes, Actinobacteria, Chlamydiae and Cyanobacteria were aggregated and the resulting proportion for each phylum plotted (blue bars). In a similar manner, the proportion of genome sequences represented by each of the indicated phyla within the TEA-MD is plotted for comparison (gray bars).
(PDF)

Figure S9 *In vitro* growth kinetics of F11 mutants. (A) F11Δ*3256*, (B) F11Δ*3082*, (C) F11Δ*2628* and (D) F11Δ*3933* were grown in M9 minimal media shaking at the indicated temperatures. Optical density (y-axis) of the cultures was recorded over time (x-axis). Graphs are representative of at least three independent experiments performed in quadruplicate.
(PDF)

Protocol S1 Description of techniques carried out during this study.
(PDF)

Table S1 General sequence-based features of input and output pools.
(PDF)

Table S2 Insertion events per 100 bp for annotated regions within the F11 chromosome.
(XLSX)

Table S3 Listing of bacterial strains, accession numbers and traits maintained within the TEA-metaproteome database.
(XLSX)

Table S4 Composition of TEA-defined pathogen-associated genes within Tn-seq gene sets.
(PDF)

Table S5 Bacterial strains and plasmids.
(PDF)

Table S6 Primers used in this study.
(PDF)

Table S7 Raw calculated metrics used to rank fitness contributions for all annotated regions within F11 by Tn-seq.
(XLSX)

Table S8 TEA-assigned trait identities for known bacterial virulence factors and conserved genes involved in basic bacterial physiology.
(PDF)

Acknowledgments

We thank Julio C. Menendez (University of Utah) for his part in developing the online data viewer. We also thank Andrew L. Goodman (Washington University School of Medicine) and Jeffrey I. Gordon (Washington University School of Medicine) for sharing the pSAM_Bt vector and bacterial strain EcS17.

Author Contributions

Conceived and designed the experiments: TJW KFF MAM. Performed the experiments: TJW JPN CWR. Analyzed the data: TJW KFF MAM. Contributed reagents/materials/analysis tools: TJW KFF BKD MAM. Wrote the paper: TJW KFF MAM.

References

1. Ewers C, Janssen T, Wieler LH (2003) [Avian pathogenic Escherichia coli (APEC)]. Berliner und Munchener tierarztliche Wochenschrift 116: 381–395.
2. Shpigel NY, Elazar S, Rosenshine I (2008) Mammary pathogenic Escherichia coli. Current opinion in microbiology 11: 60–65.
3. Tan C, Xu Z, Zheng H, Liu W, Tang X, et al. (2011) Genome sequence of a porcine extraintestinal pathogenic Escherichia coli strain. Journal of bacteriology 193: 5038.
4. Carvallo FR, Debroy C, Baeza E, Hinckley L, Gilbert K, et al. (2010) Necrotizing pneumonia and pleuritis associated with extraintestinal pathogenic Escherichia coli in a tiger (Panthera tigris) cub. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 22: 136–140.
5. Foxman B (2010) The epidemiology of urinary tract infection. Nature reviews Urology 7: 653–660.
6. Johnson JR, Johnston B, Clabots C, Kuskowski MA, Castanheira M (2010) Escherichia coli sequence type ST131 as the major cause of serious multidrug-resistant E. coli infections in the United States. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 51: 286–294.
7. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, et al. (2009) Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS genetics 5: e1000344.
8. Wiles TJ, Kulesus RR, Mulvey MA (2008) Origins and virulence mechanisms of uropathogenic Escherichia coli. Experimental and molecular pathology 85: 11–19.

9. Muzzi A, Masignani V, Rappuoli R (2007) The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug discovery today 12: 429–439.
10. Wiles TJ, Bower JM, Redd MJ, Mulvey MA (2009) Use of zebrafish to probe the divergent virulence potentials and toxin requirements of extraintestinal pathogenic Escherichia coli. PLoS pathogens 5: e1000697.
11. Wiles TJ, Norton JP, Smith SN, Lewis AJ, Mobley HL, et al. (2013) A phyletically rare gene promotes the niche-specific fitness of an E. coli pathogen during bacteremia. PLoS pathogens 9: e1003175.
12. van Opijnen T, Camilli A (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome research 22: 2541–2551.
13. Perez JC, Kumamoto CA, Johnson AD (2013) Candida albicans Commensalism and Pathogenicity Are Intertwined Traits Directed by a Tightly Knit Transcriptional Regulatory Circuit. PLoS biology 11: e1001510.
14. van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nature methods 6: 767–772.
15. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, et al. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell host & microbe 6: 279–289.
16. Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, et al. (2006) Characterization of mycobacterial virulence genes through genetic interaction mapping. Proceedings of the National Academy of Sciences of the United States of America 103: 11760–11765.
17. Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E, et al. (1995) Simultaneous identification of bacterial virulence genes by negative selection. Science 269: 400–403.
18. Saenz HL, Dehio C (2005) Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. Current opinion in microbiology 8: 612–619.
19. Hamer L, DeZwaan TM, Montenegro-Chamorro MV, Frank SA, Hamer JE (2001) Recent advances in large-scale transposon mutagenesis. Current opinion in chemical biology 5: 67–73.
20. Lehoux DE, Sanschagrin F, Levesque RC (2001) Discovering essential and infection-related genes. Current opinion in microbiology 4: 515–519.
21. Hayes F (2003) Transposon-based strategies for microbial functional genomics and proteomics. Annual review of genetics 37: 3–29.
22. Trede NS, Langenau DM, Traver D, Look AT, Zon LI (2004) The use of zebrafish to understand immunity. Immunity 20: 367–379.
23. Jault C, Pichon L, Chluba J (2004) Toll-like receptor gene family and TIR-domain adapters in Danio rerio. Molecular immunology 40: 759–771.
24. Wang Z, Zhang S, Wang G, An Y (2008) Complement activity in the egg cytosol of zebrafish Danio rerio: evidence for the defense role of maternal complement components. PloS one 3: e1463.
25. Li X, Wang S, Qi J, Echtenkamp SF, Chatterjee R, et al. (2007) Zebrafish peptidoglycan recognition proteins are bactericidal amidases essential for defense against bacterial infections. Immunity 27: 518–529.
26. Lieschke GJ, Oates AC, Crowhurst MO, Ward AC, Layton JE (2001) Morphologic and functional characterization of granulocytes and macrophages in embryonic and adult zebrafish. Blood 98: 3087–3096.
27. Lampe DJ, Churchill ME, Robertson HM (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. The EMBO journal 15: 5470–5479.
28. Smith V, Botstein D, Brown PO (1995) Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. Proceedings of the National Academy of Sciences of the United States of America 92: 6479–6483.
29. Smith V, Chou KN, Lashkari D, Botstein D, Brown PO (1996) Functional analysis of the genes of yeast chromosome V by genetic footprinting. Science 274: 2069–2074.
30. Grant AJ, Farris M, Alefounder P, Williams PH, Woodward MJ, et al. (2003) Co-ordination of pathogenicity island expression by the BipA GTPase in enteropathogenic Escherichia coli (EPEC). Molecular microbiology 48: 507–521.
31. Iyoda S, Honda N, Saitoh T, Shimuta K, Terajima J, et al. (2011) Coordinate control of the locus of enterocyte effacement and enterohemolysin genes by multiple common virulence regulators in enterohemorrhagic Escherichia coli. Infection and immunity 79: 4628–4637.
32. Nagy G, Dobrindt U, Schneider G, Khan AS, Hacker J, et al. (2002) Loss of regulatory protein RfaH attenuates virulence of uropathogenic Escherichia coli. Infection and immunity 70: 4406–4413.
33. Burns SM, Hull SI (1999) Loss of resistance to ingestion and phagocytic killing by O(−) and K(−) mutants of a uropathogenic Escherichia coli O75:K5 strain. Infection and immunity 67: 3757–3762.
34. Buckles EL, Wang X, Lane MC, Lockatell CV, Johnson DE, et al. (2009) Role of the K2 capsule in Escherichia coli urinary tract infection and serum resistance. The Journal of infectious diseases 199: 1689–1697.
35. deLivron MA, Robinson VL (2008) Salmonella enterica serovar Typhimurium BipA exhibits two distinct ribosome binding modes. Journal of bacteriology 190: 5944–5952.
36. Kiss E, Huguet T, Poinsot V, Batut J (2004) The typA gene is required for stress adaptation as well as for symbiosis of Sinorhizobium meliloti 1021 with certain Medicago truncatula lines. Molecular plant-microbe interactions : MPMI 17: 235–244.
37. Gauger EJ, Leatham MP, Mercado-Lubo R, Laux DC, Conway T, et al. (2007) Role of motility and the flhDC Operon in Escherichia coli MG1655 colonization of the mouse intestine. Infection and immunity 75: 3315–3324.
38. Morris JJ, Lenski RE, Zinser ER (2012) The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. mBio 3: e00036–12 .
39. Falkow S (2004) Molecular Koch's postulates applied to bacterial pathogenicity–a personal recollection 15 years later. Nature reviews Microbiology 2: 67–72.
40. Falkow S (1988) Molecular Koch's postulates applied to microbial pathogenicity. Reviews of infectious diseases 10 Suppl 2: S274–276.
41. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25: 3389–3402.
42. del Castillo I, Gonzalez-Pastor JE, San Millan JL, Moreno F (1991) Nucleotide sequence of the Escherichia coli regulatory gene mprA and construction and characterization of mprA-deficient mutants. Journal of bacteriology 173: 3924–3929.
43. Lomovskaya O, Lewis K, Matin A (1995) EmrR is a negative regulator of the Escherichia coli multidrug resistance pump EmrAB. Journal of bacteriology 177: 2328–2334.
44. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols 4: 363–371.
45. Arnadottir H, Hvanndal I andresdottir V, Burr SE, Frey J, et al. (2009) The AsaP1 peptidase of Aeromonas salmonicida subsp. achromogenes is a highly conserved deuterolysin metalloprotease (family M35) and a major virulence factor. Journal of bacteriology 191: 403–410.
46. Tadesse S, Graumann PL (2007) DprA/Smf protein localizes at the DNA uptake machinery in competent Bacillus subtilis cells. BMC microbiology 7: 105.
47. Karudapuram S, Zhao X, Barcak GJ (1995) DNA sequence and characterization of Haemophilus influenzae dprA+, a gene required for chromosomal but not plasmid DNA transformation. Journal of bacteriology 177: 3235–3240.
48. Smeets LC, Becker SC, Barcak GJ, Vandenbroucke-Grauls CM, Bitter W, et al. (2006) Functional characterization of the competence protein DprA/Smf in Escherichia coli. FEMS Microbiology Letters 263: 223–228.
49. Norton JP, Mulvey MA (2012) Toxin-Antitoxin Systems Are Important for Niche-Specific Colonization and Stress Resistance of Uropathogenic Escherichia coli. PLoS pathogens 8: e1002954.
50. Smith SN, Hagan EC, Lane MC, Mobley HL (2010) Dissemination and systemic colonization of uropathogenic Escherichia coli in a murine model of bacteremia. mBio 1: e00262–10.
51. Welch RA, Dellinger EP, Minshew B, Falkow S (1981) Haemolysin contributes to virulence of extra-intestinal E. coli infections. Nature 294: 665–667.
52. Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N, et al. (1999) The link between phylogeny and virulence in Escherichia coli extraintestinal infection. Infection and immunity 67: 546–553.
53. Pfennig PL, Flower AM (2001) BipA is required for growth of Escherichia coi K12 at low temperature. Molecular genetics and genomics : MGG 266: 313–317.
54. Krin E, Danchin A, Soutourina O (2010) Decrypting the H-NS-dependent regulatory cascade of acid stress resistance in Escherichia coli. BMC microbiology 10: 273.
55. Richards GR, Herbert EE, Park Y, Goodrich-Blair H (2008) Xenorhabdus nematophila lrhA is necessary for motility, lipase activity, toxin expression and virulence in Manduca sexta insects. Journal of bacteriology 190: 4870–4879.
56. Himpsl SD, Lockatell CV, Hebel JR, Johnson DE, Mobley HL (2008) Identification of virulence determinants in uropathogenic Proteus mirabilis using signature-tagged mutagenesis. Journal of medical microbiology 57: 1068–1078.
57. Whitfield C (2006) Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. Annual review of biochemistry 75: 39–68.
58. Price PA, Jin J, Goldman WE (2012) Pulmonary infection by Yersinia pestis rapidly establishes a permissive environment for microbial proliferation. Proceedings of the National Academy of Sciences of the United States of America 109: 3083–3088.
59. Hagan EC, Lloyd AL, Rasko DA, Faerber GJ, Mobley HL (2010) Escherichia coli global gene expression in urine from women with urinary tract infection. PLoS pathogens 6: e1001187.
60. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America 97: 6640–6645.
61. Murphy KC, Campellone KG (2003) Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic E. coli. BMC molecular biology 4: 11.
62. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10: R25.
63. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society, Series B 57: 289–300.
64. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28: 27–30.
65. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research 40: D109–114.

66. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic acids research 35: W182–185.
67. Rakeman JL, Bonifield HR, Miller SI (1999) A HilA-independent pathway to Salmonella typhimurium invasion gene transcription. Journal of bacteriology 181: 3096–3104.
68. Jakomin M, Chessa D, Baumler AJ, Casadesus J (2008) Regulation of the Salmonella enterica std fimbrial operon by DNA adenine methylation, SeqA and HdfR. Journal of bacteriology 190: 7406–7413.
69. Ko M, Park C (2000) H-NS-Dependent regulation of flagellar synthesis is mediated by a LysR family protein. Journal of bacteriology 182: 4670–4672.
70. Pavelka MS, Jr., Wright LF, Silver RP (1991) Identification of two genes, kpsM and kpsT, in region 3 of the polysialic acid gene cluster of Escherichia coli K1. Journal of bacteriology 173: 4603–4610.
71. Jakobsen L, Spangholm DJ, Pedersen K, Jensen LB, Emborg HD, et al. (2010) Broiler chickens, broiler chicken meat, pigs and pork as sources of ExPEC related virulence genes and resistance in Escherichia coli isolates from community-dwelling humans and UTI patients. International journal of food microbiology 142: 264–272.
72. Rigg GP, Barrett B, Roberts IS (1998) The localization of KpsC, S and T and KfiA, C and D proteins involved in the biosynthesis of the Escherichia coli K5 capsular polysaccharide: evidence for a membrane-bound complex. Microbiology 144 (Pt 10): 2905–2914.
73. Bronner D, Sieberth V, Pazzani C, Roberts IS, Boulnois GJ, et al. (1993) Expression of the capsular K5 polysaccharide of Escherichia coli: biochemical and electron microscopic analyses of mutants with defects in region 1 of the K5 gene cluster. Journal of bacteriology 175: 5984–5992.
74. McNulty C, Thompson J, Barrett B, Lord L andersen C, et al. (2006) The cell surface expression of group 2 capsular polysaccharides in Escherichia coli: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. Molecular microbiology 59: 907–922.
75. Bachtiar BM, Coloe PJ, Fry BN (2007) Knockout mutagenesis of the kpsE gene of Campylobacter jejuni 81116 and its involvement in bacterium-host interactions. FEMS immunology and medical microbiology 49: 149–154.
76. Cieslewicz M, Vimr E (1997) Reduced polysialic acid capsule expression in Escherichia coli K1 mutants with chromosomal defects in kpsF. Molecular microbiology 26: 237–249.
77. Kalynych S, Ruan X, Valvano MA, Cygler M (2011) Structure-guided investigation of lipopolysaccharide O-antigen chain length regulators reveals regions critical for modal length control. Journal of bacteriology 193: 3710–3721.
78. Amos MR, Sanchez-Contreras M, Jackson RW, Munoz-Berbel X, Ciche TA, et al. (2011) Influence of the Photorhabdus luminescens phosphomannose isomerase gene, manA, on mannose utilization, exopolysaccharide structure and biofilm formation. Applied and environmental microbiology 77: 776–785.
79. Ouyang Z, Isaacson R (2006) Identification and characterization of a novel ABC iron transport system, fit, in Escherichia coli. Infection and immunity 74: 6949–6956.
80. Stojiljkovic I, Baumler AJ, Heffron F (1995) Ethanolamine utilization in Salmonella typhimurium: nucleotide sequence, protein expression and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. Journal of bacteriology 177: 1357–1366.
81. Buan NR, Suh SJ, Escalante-Semerena JC (2004) The eutT gene of Salmonella enterica Encodes an oxygen-labile, metal-containing ATP:corrinoid adenosyltransferase enzyme. Journal of bacteriology 186: 5708–5714.
82. Turner RJ, Taylor DE, Weiner JH (1997) Expression of Escherichia coli TehA gives resistance to antiseptics and disinfectants similar to that conferred by multidrug resistance efflux pumps. Antimicrobial agents and chemotherapy 41: 440–444.
83. Nataro JP, Seriwatana J, Fasano A, Maneval DR, Guers LD, et al. (1995) Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive Escherichia coli and Shigella strains. Infection and immunity 63: 4721–4728.
84. Schwan WR (2009) Survival of uropathogenic Escherichia coli in the murine urinary tract is dependent on OmpR. Microbiology 155: 1832–1839.
85. Ware D, Jiang Y, Lin W, Swiatlo E (2006) Involvement of potD in Streptococcus pneumoniae polyamine transport and pathogenesis. Infection and immunity 74: 352–361.

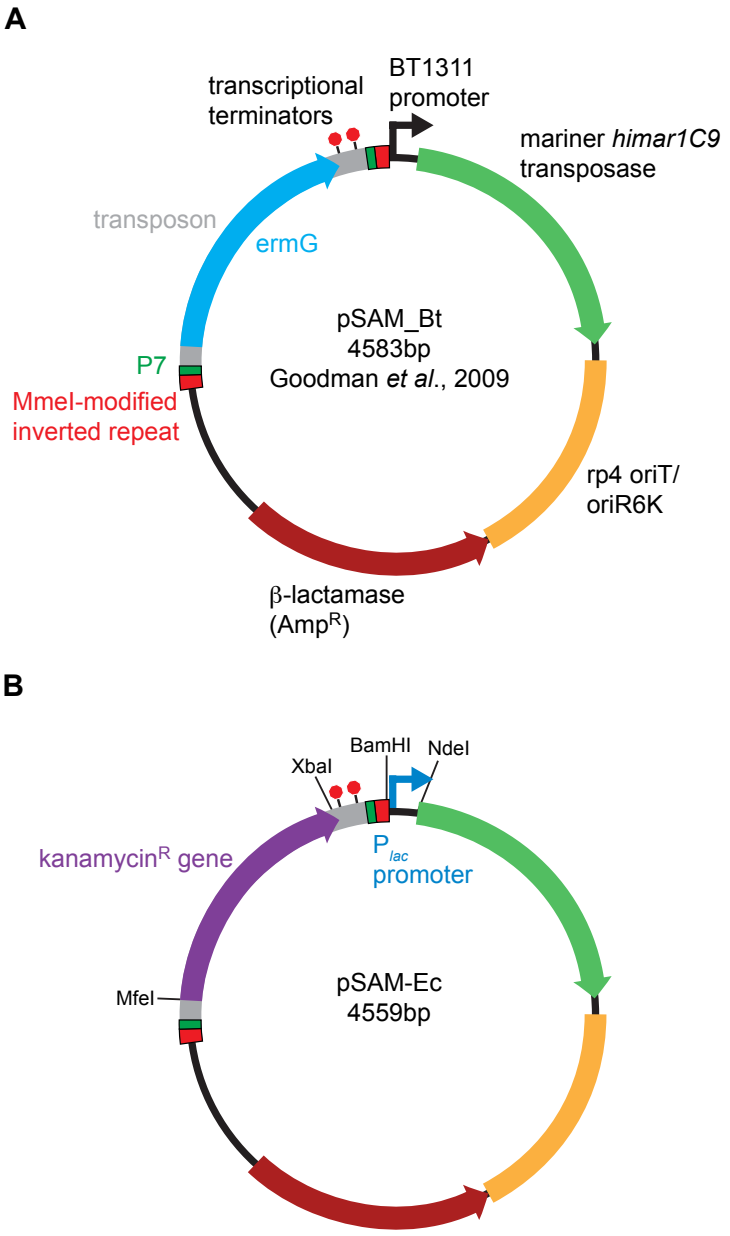


Figure A1. Plasmids used in this study. (A) pSAM_Bt was originally constructed to express the *himar1C9* transposase (light green open reading frame) within *Bacteroides thetaiotaomicron*, facilitating insertion of the *ermG* erythromycin resistance gene (blue open reading frame) into the chromosome this bacterium. (B) pSAM_Bt was retrofitted using the indicated restriction sites with a kanamycin resistance gene (purple open reading frame) and a P_{lac} promoter to drive expression of the *himar1C9* transposase within *E. coli*. This plasmid, pSAM-Ec, was then transferred into *E. coli* strain F11 via conjugation. Other pSAM-Ec features include: two transcriptional terminators downstream of the kanamycin resistance gene (red stop signs), P7 priming sites for Illumina sequencing (dark green blocks), MmeI-modified restriction sites (red blocks) flanking the transposon (grey), a β-lactamase gene for donor selection (red open reading frame) and an RP4 oriT/oriR6K origin of replication (orange block) for specific propagation in *pir*-positive donor strains.

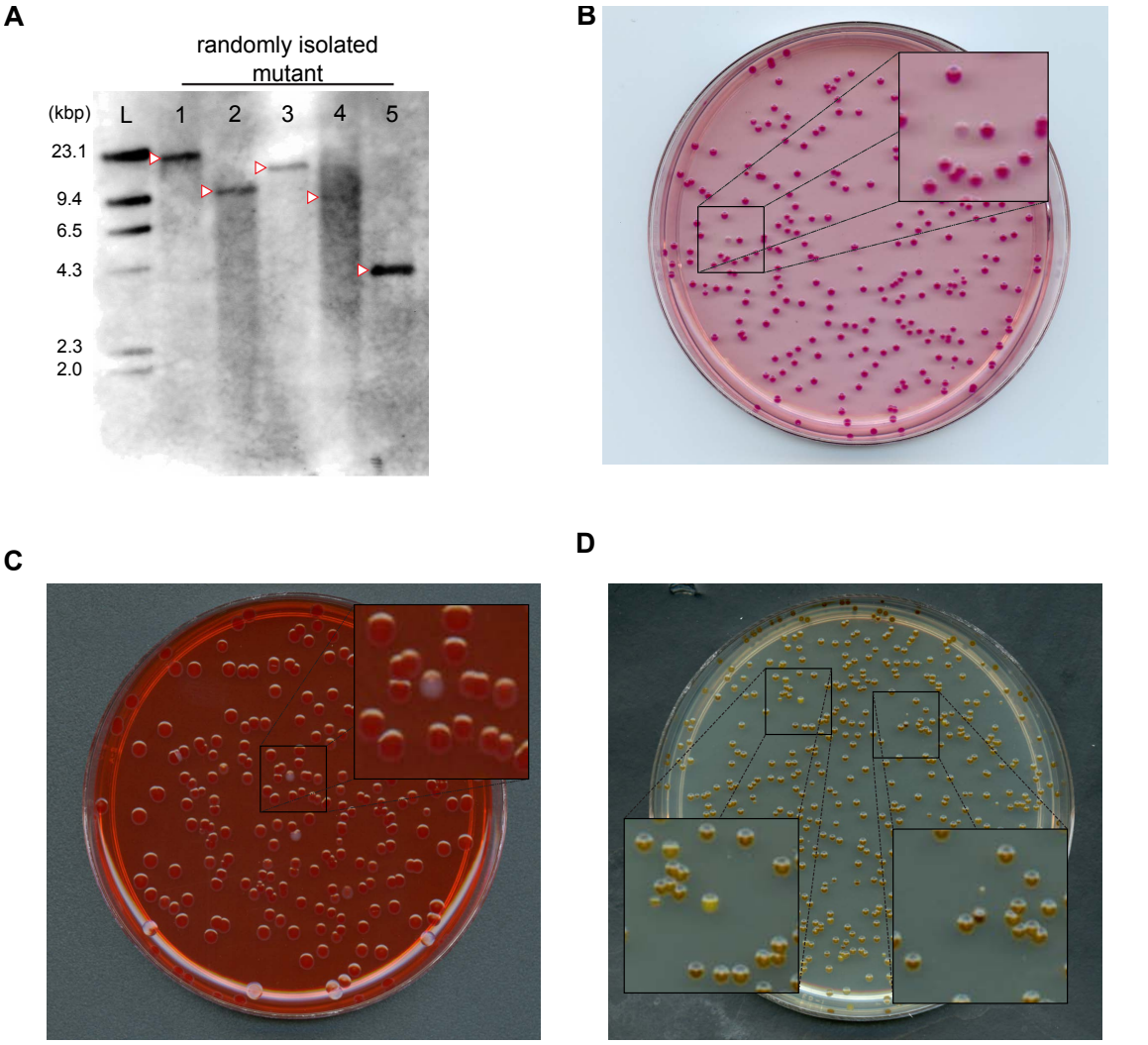


Figure A2. Assessment of transposon mutant pool diversity and saturation. (A) *E. coli* F11 transposon mutants were randomly isolated and subjected to Southern blot analysis. Genomic DNA was isolated, digested with the restriction enzyme HindIII, resolved by gel electrophoresis and probed with a digoxigenin-labeled probe specific for the kanamycin resistance gene within the transposon. (B to D) Representative colorimetric microbiological agar plates were used to estimate mutant occurrence frequency within mutant pools. (B) *E. coli* F11 transposon mutants grown on MacConkey agar to assay for the presence of mutants deficient for lactose utilization (white colony within inset). (C) Agar plates containing the dye Congo red were used to determine the frequency of insertion variants that disrupted normal curli production. Inset shows a white colony (not able to bind Congo red due to the presumed absence of curli) surrounded by curli producing mutant variants (red colonies). (D) Mutant colonies grown on Kornberg agar and subsequently exposed to iodine vapor to stain for the presence of glycogen. Normal colonies exhibit a brown coloration, colonies lacking glycogen synthesis are yellow-white (left inset) and colonies overloaded with glycogen are dark brown or black (right inset).

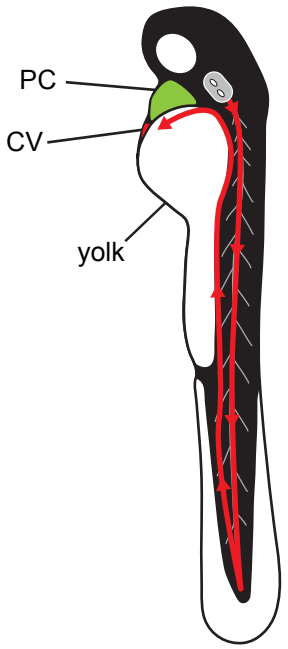


Figure A3. Diagram of a 48 h post-fertilization zebrafish embryo. *E. coli* F11 transposon mutants were delivered into zebrafish embryos via one of two injections sites. The pericardial cavity (PC) simulates a localized infection—bacteria are restricted to the indicated area (green). The circulation valley (CV) is used to deliver bacteria systemically throughout the bloodstream (red).

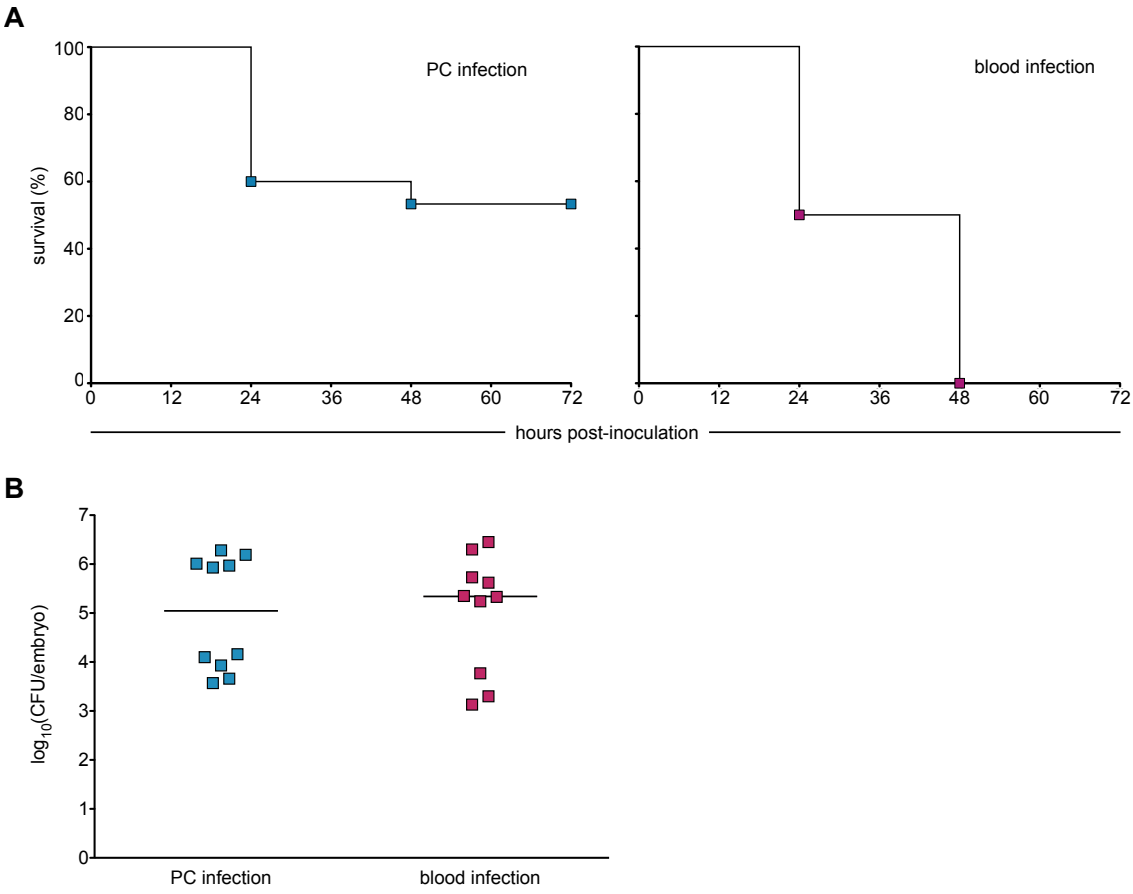


Figure A4. Virulence of *E. coli* F11 within zebrafish embryos. (A) Approximately 2,000 colony-forming units of *E. coli* F11 were delivered into the pericardial cavity (PC, left) or the blood (right). Survival of embryos was monitored over a three-day period ($n = 20$ embryos for each curve). (B) At 21 h post-inoculation, infected zebrafish were homogenized and bacterial titers determined. Each square symbol represents an individual zebrafish and bars mark median values ($n = 10$ embryos).

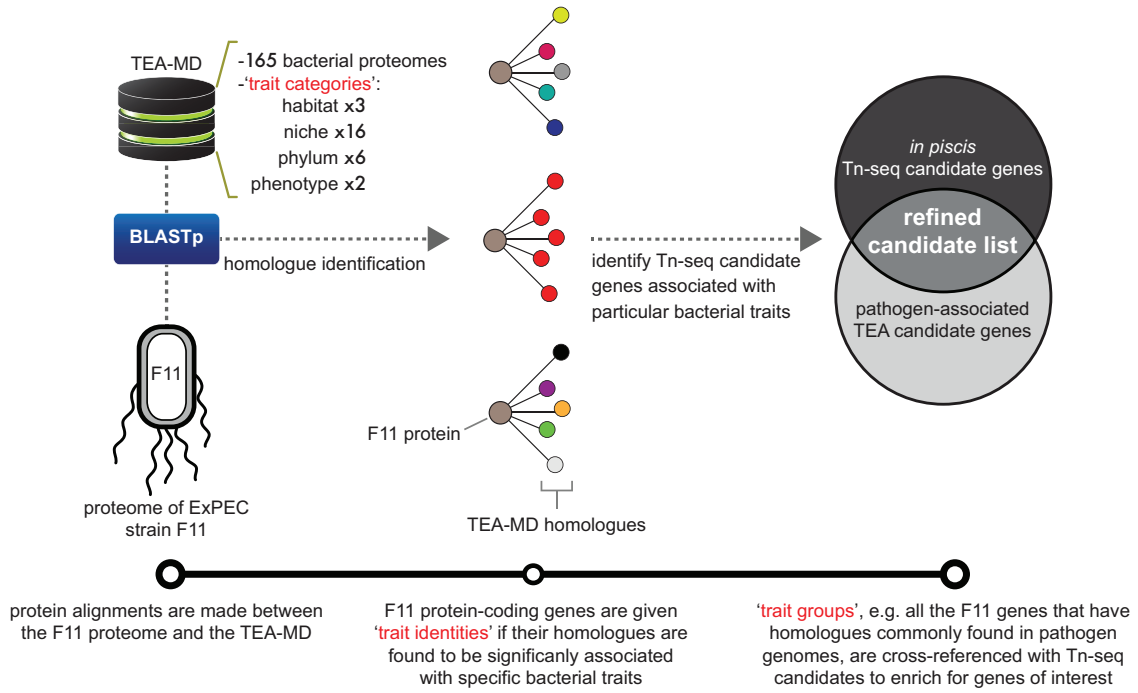


Figure A5. Description of pipeline used to define Tn-seq candidate gene sets. Diagram describes each of the five steps in determining Tn-seq-derived candidate genes and gene sets. (I) normalize insert read counts, (II) identify strong negative selection events and median-center output data, (III) perform Wilcoxon signed-rank test, (IV) initial candidate list vetting and (V) final gene set construction and curation.

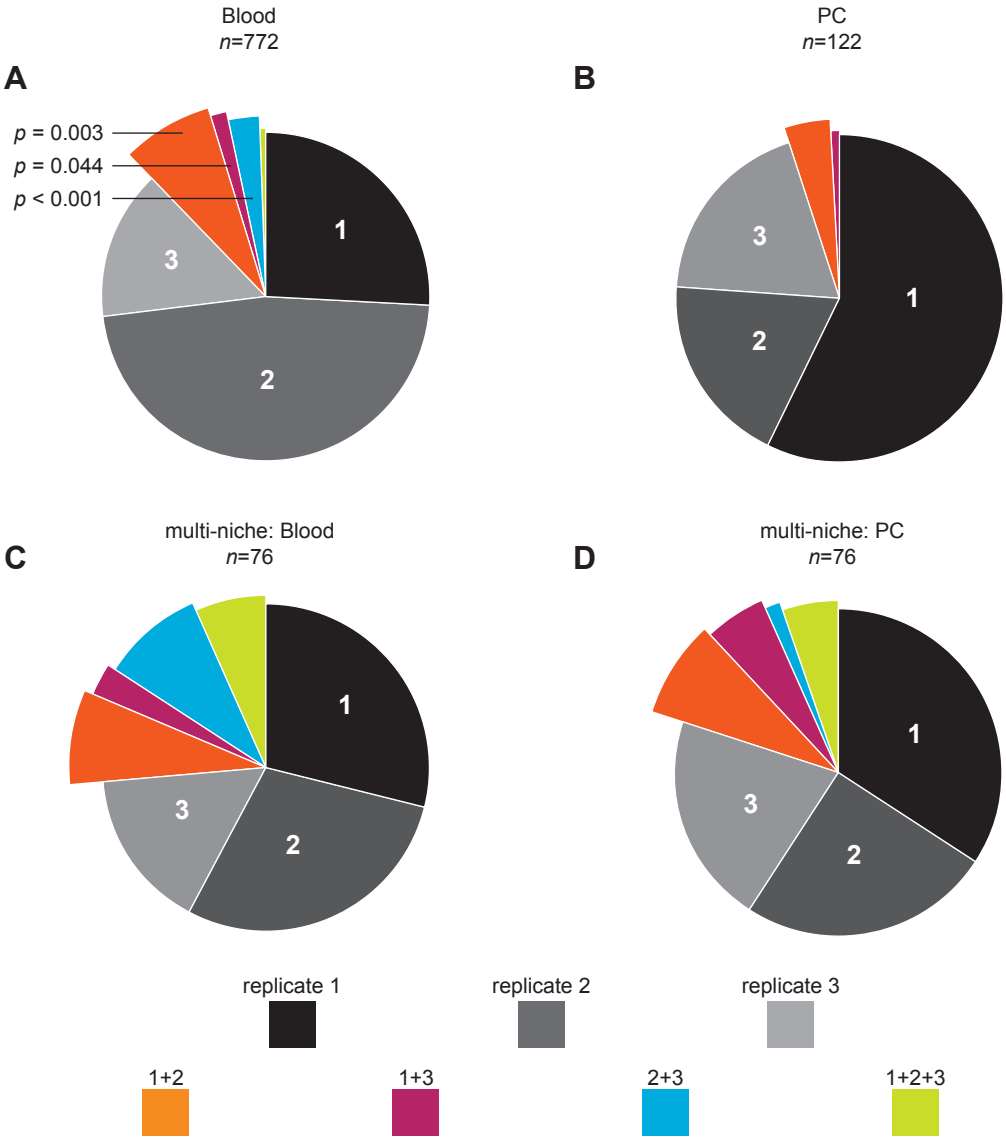
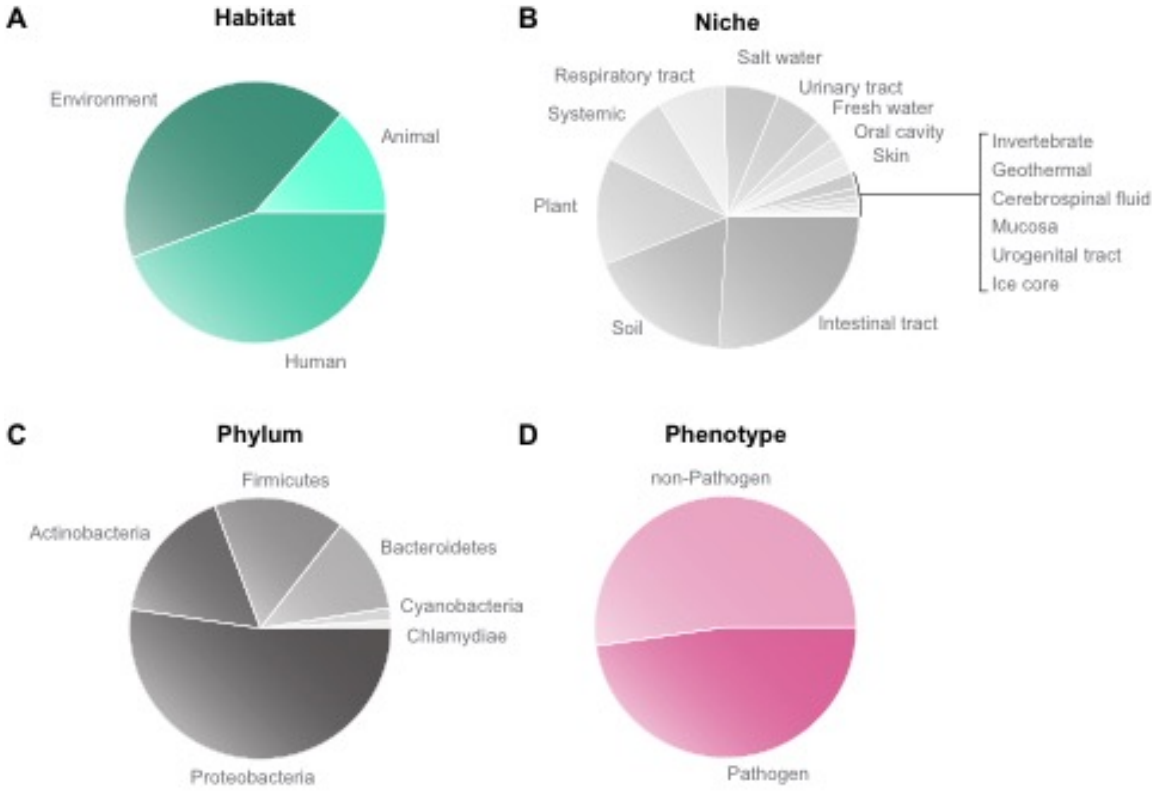


Figure A6. Relative contributions made by each replicate screen to the *in piscis* gene set. For clarity, the *in piscis* gene set was divided into the component genes sets (A) blood, (B) PC and (C and D) multi-niche. The proportion of genes identified from only one replicate screen (numbered 1, 2 and 3) is represented by dark wedges, whereas colored wedges indicate genes that were identified from two or more screens using Tn-seq (bottom key). Fischer's exact test was used to determine if there was significant overlap between genes contributed by each replicate screen within each gene set (*p* values indicated where significance was found).



E

Habitat	total genomes	total proteins
Animal	26	83,103
Environment	59	258,787
Human	80	272,437
Niche		
Ice core	1	2,534
Urogenital tract	2	3,172
Mucosa	4	4,753
Cerebrospinal fluid	1	4,827
Geothermal	2	6,178
Invertebrate	5	12,071
Skin	4	13,236
Oral cavity	8	15,903
Fresh water	4	16,455
Urinary tract	8	35,554
Salt Water	12	41,606
Respiratory tract	14	50,777
Systemic	19	55,106
Plant	17	81,255

Niche *continued*	total genomes	total proteins
Soil	23	110,759
Intestinal tract	41	160,141
Phylum		
Chlamydiae	5	5,822
Cyanobacteria	4	9,596
Bacteroidetes	22	73,311
Firmicutes	30	99,767
Actinobacteria	26	104,210
Proteobacteria	78	321,621
Phenotype		
non-Pathogen	81	321,251
Pathogen	84	293,076

Figure A7. Compositional features of the TEA metaproteome database (TEA-MD). Pie graphs depict the proportion of protein sequences contained within the TEA-MD that are contributed by bacteria annotated with the indicated traits under the trait categories (A) habitat of isolation, (B) niche of isolation, (C) phylum and (D) phenotype. (E) Table detailing the total number of bacterial genomes and respective aggregate of protein sequences represented by specific traits within each trait category.

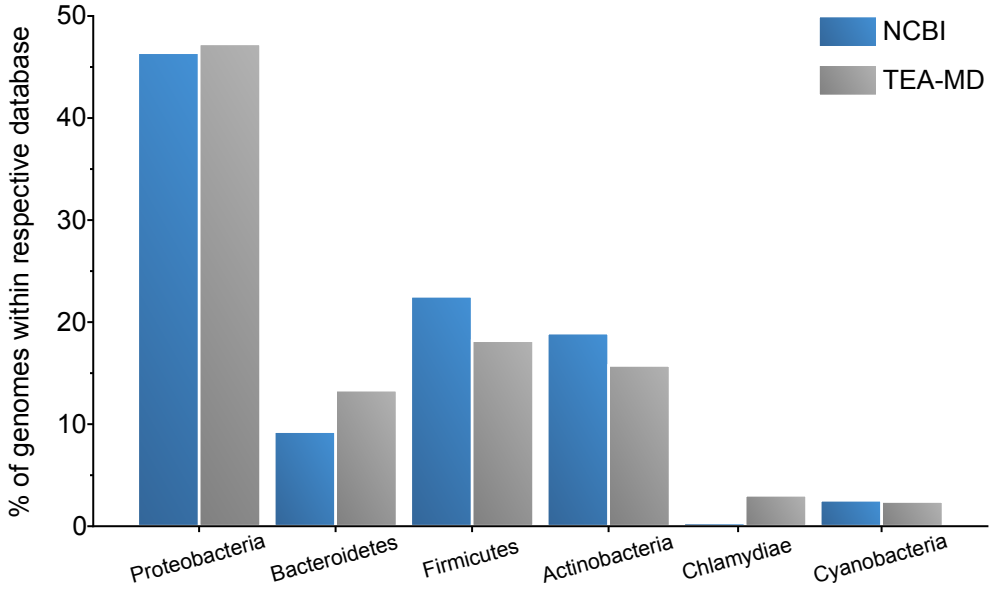


Figure A8. Comparison of phylum representation between NCBI and the TEA-MD. The number of genome sequences currently reported in the National Center for Biotechnology Information (NCBI) database for Proteobacteria, Bacteroidetes, Firmicutes, Actinobacteria, Chlamydiae and Cyanobacteria were aggregated and the resulting proportion for each phylum plotted (blue bars). In a similar manner, the proportion of genome sequences represented by each of the indicated phyla within the TEA-MD is plotted for comparison (gray bars).

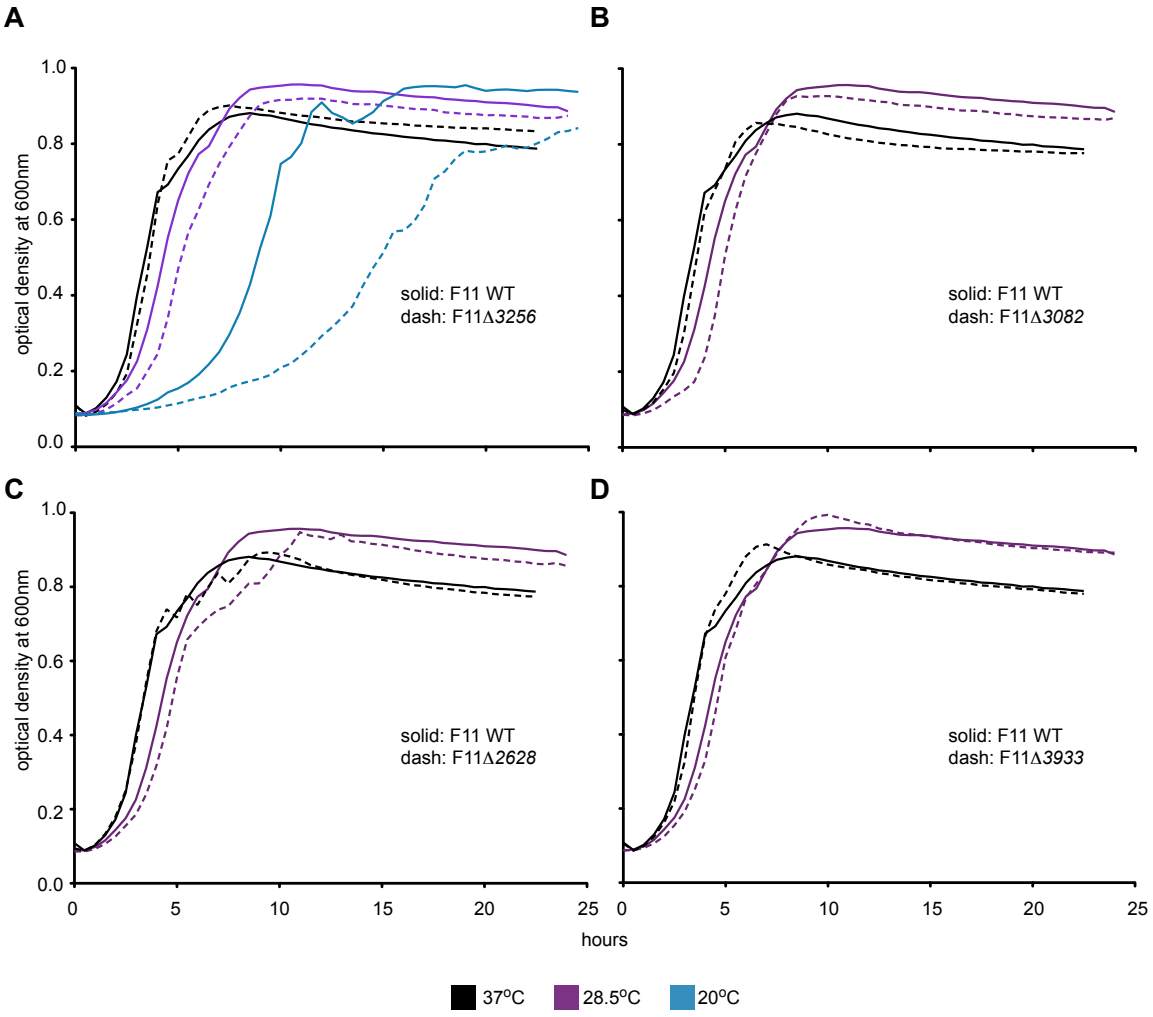


Figure A9. *In vitro* growth kinetics of F11 mutants. (A) F11Δ*3256*, (B) F11Δ*3082*, (C) F11Δ*2628* and (D) F11Δ*3933* were grown in M9 minimal media shaking at the indicated temperatures. Optical density (y-axis) of the cultures was recorded over time (x-axis). Graphs are representative of at least three independent experiments performed in quadruplicate.